

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 41 Seconds

(without alignments)
553,556 Million cell updates/sec

Title: US-09-701-675A-3

Perfect score: 1210
Sequence: 1 MLSKGLKRRKEEKEPLA.....YLMADVLTQALRPPGGR 236

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Minimum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	120	9.9	3164	1 WMBEH6	UL36 protein - hum
2	112.5	9.3	593	2 I51213	drebrin - chicken
3	109	9.0	688	3 I51249	probable virion pr
4	108.5	9.0	1217	2 T00270	hypothetical prote
5	107.5	8.9	440	2 S71795	transcription fact
6	107.5	8.9	708	1 Q0EBB8	UL80 protein - hum
7	107	8.8	1317	2 T03748	apoptosis associat
8	106.5	8.8	512	2 H84310	cobryric acid synth
9	106.5	8.8	753	2 T46614	chemotaxis protein
10	106.5	8.8	909	2 S32538	CGMP-gated cation
11	105	8.7	607	2 A43776	drebrin E2 - chick
12	102.5	8.5	801	2 T29018	hypothetical prote
13	102.5	8.5	901	2 A44825	phosphoprotein, sy
14	102.5	8.5	908	2 A33380	sarcalumenin precu
15	101.5	8.4	676	1 EDBE22	immediate-early pr
16	101.5	8.4	1095	2 T00329	hypothetical prote
17	101.5	8.4	2774	2 A43359	microtubule-associ
18	101	8.3	433	2 S69999	sterigmatocystin s
19	101	8.3	1532	2 A61262	collagen alpha 1(X
20	100.5	8.3	676	1 EDBE23	immediate-early pr
21	100	8.3	655	2 S40521	FKHR protein - hum
22	100	8.3	1851	2 T19964	hypothetical prote
23	99.5	8.2	395	2 I49575	CCAAT/enhancer bin
24	99.5	8.2	1106	1 A39299	Bassoon protein -
25	99.5	8.2	3942	1 T42730	secreted glycoprot
26	98.5	8.1	498	1 VGBEGX	collagen, cornea-s
27	98.5	8.1	1446	2 A38587	bullous pemphigoid
28	98.5	8.1	1433	2 A46053	hypothetical prote
29	98	8.1	337	2 T21055	

30	98	8.1	1199	2 A40670
31	98	8.1	2282	1 T42717
32	98	8.1	3010	1 GNMVTV
33	96.5	8.0	590	1 T35297
34	96.5	8.0	1207	2 T00378
35	96.5	8.0	1315	2 A56101
36	96.5	8.0	1774	2 B56101
37	96	7.9	416	1 SKXLAG
38	96	7.9	477	2 T46304
39	96	7.9	936	2 D97630
40	96	7.9	1298	1 EDBE75
41	95.5	7.9	245	1 W4WU5
42	95.5	7.9	303	1 TVH0JD
43	95.5	7.9	614	2 T33149
44	95.5	7.9	672	2 I40333
45	95.5	7.9	1257	2 S28764

nuclear envelope p
DNA-binding protei
genome polyprotein
probable dihydrol
KIAA0641 protein
collagen alpha 1(X
dermal gland prote
hypothetical prote
hypothetical prote
immediate-early pr
E4 protein - human
transforming prote
hypothetical prote
tracheal colonizat
neurocan precursor

ALIGNMENTS

RESULT 1

WMBEH6
UL36 protein - human herpesvirus 1 (strain 17)

C:Species: human herpesvirus 1

C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000

C:Accession: I30085

R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; J

J. Gen. Virol. 69, 1531-1574, 1988

A:Title: The complete DNA sequence of the long unique region in the genome of herpes

A:Reference number: A30083; M01D:88274327; PMID:2839594

A:Accession: I30085

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-3164 <MGCS>

A:Cross-references: GB:X14112; NID:91944536; PIDN:CAA3231.1; PID:959536; GB:D00317

A:Gene: UL36

A:Superfamily: varicella-zoster virus gene 22 protein

Query Match 9.9%; Score 120; DB 1; Length 3164;
Best Local Similarity 26.7%; Pred. No. 2.8;
Matches 65; Conservative 23; Mismatches 107; Indels 48; Gaps 10;

QY	21	VDSWVDPGHAVAQAAPVASSLSFEDSVLKLHSHLSQSEPDRLHLYVNTLRIGQS	80
DB	2396	ITWAVRLHADLNPTENACIAAOLPRISALTAERPLARGP-----DIS	2445
QY	81	MAPAAALPPVSPAPASV-----ADNLLASS-----DAATASNASLLE	120
DB	2446	MPPVAVLWENPDPPGPPDVRFGSEATELPPVAGGEDVLASATDEDPFLARALIGRF	2505
QY	121	DLSHIEGSLQAPLADSEPPGSRISGGAAPSIGALD-----LGPAT--GCLLDGLEG	172
DB	2506	DASLSGELFPGHVYQARAPDQSSVNPPTGPPVDLVGABESLPGSLATLFLFDAPRG	2565
QY	173	LEPDITSYDNEELNAPASEGLKPGDEDPGKEAPELDEAEVLDMVLGT--QALERP	231
DB	2566	--EPVPPRMW---ANIHGLEELASDSCGP-----APLAPPLSPTRADQSPVTSQCARR	2616
QY	233	PGP 234	
DB	2617	PGP 2619	

SEQUENCE COMPARISON A

RESULT 2
151213
drebrin - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 21-Jul-2000
R:Kojima, N.; Shitao, T.; Odate, K.
Brain Res. Mol. Brain Res. 19, 101-114, 1993

THIS PAGE BLANK (USPTO)

A:Title: Molecular cloning of a developmentally regulated brain protein, chicken drebrin
A:Reference number: 151212; MWID:93368392; PMID:8361332
A:Accession: 151213
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-593 <KOJ>
A:Cross-references: GB:S65296; NID:g410604; PIDN:AAB28012.1; PID:g410605
C:Genetics:
A:Introns: 26/3; 51/3; 100/3; 126/3; 177/2; 198/3; 218/3; 257/1; 303/1; 346/1; 536/3; 57

Query Match 9.3%; Score 112.5; DB 2; Length 593;
Best Local Similarity 25.2%; Pred. No. 1.3;
Matches 63; Conservative 18; Mismatches 76; Indels 93; Gaps 13;

QY 5 GLKRRKEEEE-----KEPLAVDSWMLDPCHAVAQAPPAVASSSLFDLS 49
DB 329 GEORTRAPAEETPATKSPSPSTOYAEPAATEOHPPFPEDKAEPPG----- 378
50 VLKLSHSLQOSEPD-----LRHLVLYVNTLRRIQASMAPAALPPVSPPAAPS 98
379 -----DEPPDPAPAPATAGADVIGDLY-----TLPESEPPAPAAAS--EPQVETPG 423
99 VADNL--LASSDAALSASMSALLEDSHIEGLSQAPQPLADEGPPGSGISGAAPSLGALD 156
DB 424 VAEPLLELMQSDGAAPAA-----TSTWPLPDPAPAGPP-----VPEEGT-- 462
QY 157 LIG-----PARGCLLDGDELFEDIDTSMDNELMA-----PASEGLKPGEDDGRKE 205
DB 463 LIGLDELPEPPAFCDQAEQ-----HEVEEEEEETATGEPHPTGLGVOEGVOEGP--- 514
QY 206 EAEPLDEAEL 215
DB 515 EEPPTNGEM 524

RESULT 3

S61249
Probable v1rion protein - bovine herpesvirus 1
C:Species: bovine herpesvirus 1
C>Date: 18-Sep-1997 #sequence_revision 18-Sep-1997 #text_change 26-Aug-1999
C:Accession: S61249
R:Vicek, C.; Benes, V.; Lu, Z.; Kutish, G.F.; Paces, V.; Rock, D.; Letchworth, G.J.; Sch
submitted to the EMBL Data Library, January 1995
A:Description: Nucleotide sequence analysis of a 30-kb region of the bovine herpesvirus
A:Reference number: S61233
A:Accession: S61249
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-688 <VIC>
A:Cross-references: EMBL:248053; NID:g971311; PIDN:CAA8127.1; PID:g971328
Superfamily: varicella-zoster virus gene 54 protein

Query Match 9.0%; Score 109; DB 2; Length 688;
Best Local Similarity 26.6%; Pred. No. 2.8;
Matches 45; Conservative 25; Mismatches 47; Indels 52; Gaps 7;

QY 27 DPGHAAYAOAPPAVASSSLFDLSYLKX-----HSIQOSEPDRLHLYVNTL----- 74
DB 255 BEGCEYLVMPVWFGGYLFESEPMQRLHGEVYLACHALREH---ARICQLNTVPLKVLV 310
QY 75 -RRIOASMAPAALPPVSPPAAPSVADNLASSDAALSASMSALLEDSHIEGLSQAPQ 133
DB 311 GRRAEALRA-----KPPGAPSVTEKLIGEGEDAAASSAARLILITLYNMGM----- 356
QY 134 PLADEGPPGSGISGAAPSLGALDLGPAATGCLLDGLEGLEF--DIDTS 180
DB 357 -----RHIGDISSETVS-----YIDDTAAGLFDVSDVPTS 386

RESULT 4

S60270
Chemical protein KIAA0596 - human (fragment)
A:Residues: Homo sapiens (man)

C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00270
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A:Reference number: 214086; MWID:98290545; PMID:9628581
A:Accession: T00270
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1217 <NAG>
A:Cross-references: EMBL:AB011168; NID:g3043715; PIDN:BA25522.1; PID:g3043716
A:Experimental source: brain
C:Genetics:
A>Note: KIAA0596

Query Match 9.0%; Score 108.5; DB 2; Length 1217;
Best Local Similarity 26.2%; Pred. No. 6;
Matches 65; Conservative 21; Mismatches 91; Indels 71; Gaps 12;

QY 22 DSWMLDPCHAAYAOAPPAVASSSLFDLSYLKLSHSLQOSEPDRLHLYVNTLRRIQASM 81
DB 881 DSSW-----APKRVATASPF--SGLQAKQSVSHLVQERH-----EAL 917
QY 82 -APA-----AALPPVSPPAAPSVADNLASSDAALSASMSALLEDSHIE 125
DB 918 QAPSPGALLSRIEAGDGLSLPADGPPSRPHSYQNPTSSMAKISRST--SGVENL--- 973
QY 126 EGISQAPQ-----PLADEGPPGSGISGAAPSLGALDLGPAATGCLLDGLE- 171
DB 974 -GLVAPDQAHAPIRVPSLSKLLPSRAHLVLDIPKPLPRTLAAPSPVTKGRAPGEAKR 1032
QY 172 -GLFEDIDTSMDNELMAPEASEGLKPGP-----EDPGKEAPEDEBAELDVLMDVLG 224
DB 1033 PGFPVGLGKAHSTTERNAACIGESTTPKPRTECOAHGFPSSPCAOQLPVSSL-----FQG 1086
QY 225 TQALERPP 232
DB 1087 PENLQPPP 1094

RESULT 5

S71795
Transcription factor CBF-2 - chicken
C:Species: Gallus gallus (chicken)
C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jul-2000
C:Accession: S71795
R:Yanase, J.; Hirano, S.; Yamagata, M.; Noda, M.
Nature 382, 632-635, 1996
A:Title: Visual projection map specified by topographic expression of transcription f
A:Reference number: S71794; MWID:96338226; PMID:8757134
A:Accession: S71795
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-440 <YNA>
A:Cross-references: EMBL:U47276; NID:g1546783; PIDN:AAB08467.1; PID:g1546784
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F:143-234/Domain: fork head DNA-binding domain homology <FHN>

Query Match 8.9%; Score 107.5; DB 2; Length 440;
Best Local Similarity 32.3%; Pred. No. 2.1;
Matches 40; Conservative 12; Mismatches 45; Indels 27; Gaps 6;

QY 82 APAALPPVSPPAAPSVADNLASSDA-----LSASMSALLEDSHIEGLSQAP 132
DB 315 APAALPPPPPPPPRRAPLLPALELATPPGYPHPALAAASLHAAPGSGAAVAR 374
QY 133 QPLADEGPPGSGISGAAPSLGALDLGPAATG---CLLDG--LEGFEDIDTSMDNELM 187
DB 375 SPFSISIIIG---GGGPGGLGA---GPAAGAGSCASGAGATGSLRSLSGSL----- 421
QY 188 APAS 191
DB 422 APAA 425

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 41 seconds
(without alignments)
553.556 Million cell updates/sec

Title: US-09-701-675A-3
Perfect score: 1210
Sequence: 1 MLSKGLKRRREDEKEPLA.....YLMADYLVGTQALRRPPGGR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Minimum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR76:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	120	9.9	3164	1 WMBEH6	UL36 protein - hum
2	112.5	9.3	593	2 S1213	drebrin - chicken
3	109	9.0	688	2 S61249	probable virion pr
4	108.5	9.0	1217	2 T00270	transcription fact
5	107.5	8.9	440	2 S71795	UL80 protein - hum
6	107.5	8.9	708	1 Q08EB8	apoptosis associat
7	107	8.8	1317	2 T03748	cobrytic acid synth
8	106.5	8.8	512	2 H84310	chemotaxis protein
9	106.5	8.8	753	2 S32538	CGMP-gated cation
10	106.5	8.8	909	2 S32538	drebrin E2 - chick
11	105	8.7	607	2 A43776	phosphoprotein, sy
12	102.5	8.5	801	2 T29018	hypothetical prote
13	102.5	8.5	901	2 A44825	phosphoprotein, sy
14	102.5	8.5	908	2 A33280	immediate-early pr
15	101.5	8.4	676	1 EDBE22	microtubule-assoc
16	101.5	8.4	1095	2 T00329	sterigmatocystin s
17	101.5	8.4	2774	2 A43359	collagen alpha 1(X
18	101	8.3	433	2 S69999	immediate-early pr
19	101	8.3	1532	2 A61262	microtubule-assoc
20	100.5	8.3	676	1 EDBE23	sterigmatocystin s
21	100	8.3	655	1 EDBE23	collagen alpha 1(X
22	100	8.3	1851	2 T19964	immediate-early pr
23	99.5	8.2	395	2 T49575	hypothetical prote
24	99.5	8.2	1106	1 A39299	CCAAT/enhancer bin
25	99.5	8.2	3942	2 T42730	DNA-directed DNA p
26	99	8.2	498	1 VGBEGX	Bassoon protein -
27	98.5	8.1	1446	2 A38587	secreted glycoprot
28	98.5	8.1	1433	2 A46053	collagen, cornea-s
29	98	8.1	337	2 T21053	bullous pemphigoid
					hypothetical prote

30	98	8.1	1199	2 A40670	nuclear envelope p
31	98	8.1	2282	2 T42717	DNA-binding protei
32	98	8.1	3010	1 GNMATW	genome polyprotein
33	96.5	8.0	590	1 T35297	probable dihydrop
34	96.5	8.0	1207	2 T00378	KIAA0641 protein -
35	96.5	8.0	1315	2 A56101	collagen alpha 1(X
36	96.5	8.0	1774	2 B56101	collagen alpha 1(X
37	96	7.9	416	1 SKXLG	dermal gland prote
38	96	7.9	477	2 T46304	hypothetical prote
39	96	7.9	936	2 D97630	hypothetical prote
40	96	7.9	1298	1 EDBE75	immediate-early pr
41	95.5	7.9	245	1 W4WLS	E4 protein - human
42	95.5	7.9	303	1 TVH0UD	transforming prote
43	95.5	7.9	614	2 T33149	hypothetical prote
44	95.5	7.9	672	2 I40333	tracheal colonizat
45	95.5	7.9	1257	2 S28764	neutocan precursor

ALIGNMENTS

RESULT 1
WMBEH6
UL36 protein - human herpesvirus 1 (strain 17)
C:Species: human herpesvirus 1
C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000
C:Accession: I30085
R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; P
J. Gen. Virol. 69, 1531-1574, 1988
A:Title: The complete DNA sequence of the long unique region in the genome of herpes
A:Reference number: A30083; MID:8824327; PMID:2839594
A:Accession: I30085
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3164 <MCG>
A:Cross-references: GB:X14112; NID:q1944536; PIDN:CAA32311.1; PID:q59536; GB:D00317
A:Genes:
C:Superfamily: varicella-zoster virus gene 22 protein
Query Match 9.9%; Score 120; DB 1; Length 3164;
Best Local Similarity 26.7%; Pred. No. 2.8;
Matches 65; Conservative 23; Mismatches 107; Indels 48; Gaps 10;
QY 21 VQSWMLDPGAAVAAQAPPAVASSLSFDLVKLKHSIQSEPDRLRLVYVNTLRIOAS 80
DB 2396 IETWAVRLHADINPIENACLAQPLRLAERPLANGPP---CLVIV-----DTS 2445
QY 81 MAPAALPPVPSPPAASV-----ADNLAS8---DAALSASMSLDE 120
DB 2446 MPVPAVLMENPDPPGPPVRFVGSSEATEELPFVAGEEDVLAASATDEDFLARILGRPF 2505
QY 121 DLSHIEGLSQAPQPLADEBPPGSGISGAAPSGAIDL-----LGPAT---GCLLDGLEG 172
DB 2506 DMSLISGELFPGHPPVQARPDQSPVPNPTGPPVDLVGAEGSLGPGSLAPLFTDAPG 2565
QY 173 LPEDIDTSMYDNEIMAPASEGLKPGEDPGKGEAEPELDEALDYIMDVLYGT-QALERP 231
DB 2566 --EPVPPRW---AWIHGLEELASDSSGP---APLAPDPPLSPAPDSVPISQCAPRP 2616
QY 232 PGP 234
DB 2617 PGP 2619
RESULT 2
I51213
drebrin - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 21-Jul-2000
C:Accession: I51213
R:Koijima, N.; Shirao, T.; Obara, K.
Brain Res. Mol. Brain Res. 19, 101-114, 1993

```

A:Title: Molecular cloning of a developmentally regulated brain protein, chicken drebrin
A:Reference number: 151212; MUID:93368392; PMID:8361332
A:Accession: 151213
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-593 <KO>
A:Cross-references: GB:565296; NID:g410604; PIDN:AMB28012.1; PID:g410605
C:Genetics: 26/3; 51/3; 100/3; 126/3; 177/2; 198/3; 218/3; 257/1; 303/1; 346/1; 536/3; 57
Query Match
Best Local Similarity 9.3%; Score 112.5; DB 2; Length 593;
Matches 63; Conservative 18; Mismatches 76; Indels 93; Gaps 13;
OY 5 GKKRRKEE-----KEPLAVDSKMLDPGHAANAQAPPAVASSSLFDLS 49
Db GEORIRAPAEETPATPKDPSPTQVAEPATEQHWPPGPEDKAAEPPG----- 378
OY 329 GEORIRAPAEETPATPKDPSPTQVAEPATEQHWPPGPEDKAAEPPG----- 378
Db VLKLIHSLQGEPP-----LRILVYVWTLRLRIGASMAPAALPVPSPAPAS 98
OY 379 -----DEPDPRPRAWTAGADVLGDLV-----TLSESESPAPAS--EPVPETPG 423
Db 99 VADNL--LASSDAALASMASLLEDLSHIEGLSQAPOPLADEGPGRSIGGAPSLGALD 156
OY 424 VAEPILIELMQSDGAPAA-----TSWPLPVDTPAGPP-----VPEEGV-- 462
Db 157 LLG-----PATGLDDGLEGLEFEDIDTMYNDELMA-----PASEGLKGPEDGEGKE 205
OY 463 LLGDELPEPAPTCAAEQ-----HEVEEEEEEAATAGEPHPTGLGYOGEYGEV-- 514
Db 206 EAPELDEAEL 215
OY 515 EVPPTNGEM 524
Db
RESULT 3
S61249
probable virion protein - bovine herpesvirus 1
C:Species: bovine herpesvirus 1
C:Date: 18-Sep-1997 #sequence_revision 18-Sep-1997 #text_change 26-Aug-1999
C:Accession: S61249
R:Vleck, C.; Benes, V.; Lu, Z.; Kutish, G.F.; Paces, V.; Rock, D.; Letchworth, G.J.; Sci
submitted to the EMBL Data Library, January 1995
A:Description: Nucleotide sequence analysis of a 30-kb region of the bovine herpesvirus
A:Reference number: S61233
A:Accession: S61249
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-688 <EMBL>
A:Cross-references: EMBL:Z48053; NID:g971311; PIDN:CAA88127.1; PID:g971328
C:Superfamily: varicella-zoster virus gene 54 protein
Query Match
Best Local Similarity 9.0%; Score 109; DB 2; Length 688;
Matches 45; Conservative 25; Mismatches 47; Indels 52; Gaps 7;
OY 27 DPGHAANAQAPPAVASSSLFDLSVLKLH-----HSIQOSEPDLRLHYLVNTL----- 74
Db 255 EEGECFYLMAVWNGDVLFESEPMQRLHGEVLAChALREH---ARLCQLLNTVPLRVLY 310
OY 75 -RRIQASMAPAALPVPSPAPASVADNLLASSDAALASMASLLEDLSHIEGLSQAPQ 133
Db 311 GRRAALRA-----KPGAPSVKELLEGEDDAASSAANKILILYNNKGM----- 356
OY 134 PLADGPGRSIGGAPSLGALDILGPATGCLLDGLEGLEF--DIDTS 180
Db 357 -----RHIGDISETVRS-----YLDDTAAGLFDVSDVDT 386

```

```

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T002270
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T002270
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1217 <NMG>
A:Cross-references: EMBL:AB011168; NID:g3043715; PIDN:BAA25522.1; PID:g3043716
A:Experimental source: brain
C:Genetics:
A:Note: KIAA0596

Query Match          9.0%; Score 108.5; DB 2; Length 1217;
Best Local Similarity 26.2%; Pred. No. 6;
Matches 65; Conservative 21; Mismatches 91; Indels 71; Gaps 12;

QY 22 DSWMLDPCHAAVAQAAPRVAVASSLFDLSVLKHLHSLDQSEPDHLVLYVWTLRIOASM 81
DB 861 DSSW-----APKRVATASPF--SGLKAQSVSHLVQEH-----EASL 917
QY 82 -APA-----AALPVPSPPAPSVADNLLASDPAALSASMSALLEDSHI 125
DB 918 QAPSGGALLSREIAQDGLSPADGPPSPKSHYQNPNTSSMAKISRST-SVENL--- 973
QY 126 EGLSQAPQ-----PLADEGPPGRS-----IGAAPSLGALLDLGPATGCLLDGLE- 171
DB 974 -GLVAEPQAHAPIRVSPLSKLALPSRAHLVLDIEPKLPDRPTTLAASFVTKGRAPGAEK 1032
QY 172 -GLEPEDITSMYNDELNAPASEGLKPGP-----EDGPKGEAPELDEAELDIYMDLVG 224
DB 1033 PGFPVGLCKAHSSTERWACLGEGTTPKRTCEQAHPGPSSPCAQQLPVSSSL-----FQG 1086
QY 225 TQALERPP 232
DB 1087 PENIQPP 1094

RESULT 5
transcription factor CBF-2 - chicken
S:171795
C:Species: Gallus gallus (chicken)
C:Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jul-2000
C:Accession: S71795
R:Yusasa, J.; Hirano, S.; Yamagata, M.; Noda, M.
Nature 382, 632-635, 1996
A:Title: Visual projection map specified by topographic expression of transcription f
A:Reference number: S71794; MUID:96338226; PMID:8757134
A:Accession: S71795
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-440 <YBU>
A:Cross-references: EMBL:U47276; NID:g1546783; PIDN:AAH08467.1; PID:g1546784
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology
F:143-233/Domains: fork head DNA-binding domain homology <FHD>

Query Match          8.9%; Score 107.5; DB 2; Length 440;
Best Local Similarity 32.3%; Pred. No. 2.1;
Matches 40; Conservative 12; Mismatches 45; Indels 27; Gaps 6;

QY 82 APAAALPVPSPAPAPSVADNLLASDAA-----LSASMSALLEDSIHTEGLSQAP 132
DB 315 APAAALPVPSPPPPPRRARAPLLPAEELARPFQPHPLPALAASDHAAPGSGAAVAR 374
QY 133 QPLADEGPPGSGIGGAAPSGALLDLGPATG-----CLLDGQ-LEGLPEPEDITSMYNELW 187
DB 375 SPFSIESITIG--GGPGGLGA---GPAPGAGSCASGGAATGLSRSLGSGL----- 421
QY 188 APAS 191
DB 422 APAA 425

```

A:	Accession:	T03748
A:	Status:	preliminary; translated from GB/EMBL/DDBJ
A:	Molecule type:	mRNA
A:	Residues:	1-1317 <GAO>
A:	Cross-references:	EMBL:AF011908; NID:g2459992; PIDN:AAB71837.1; PID:g2459993
C:	Genetics:	
A:	Gene:	ATYK
Query Match	Best Local Similarity	8.8%; Score 107; DB 2; Length 1317;
Matches	58; Conservative	20; Mismatches 79; Indels 88; Gaps 10;
Oy	78 QASMAPAALPPVSP-----PAAPSTADN-----LIASSDA	109
Db	746 EGSAPDLPVPSPVSCGASLPSEBAPDLIPPTPAAGSWTVPEPAPTLES	805
Oy	110 ALSMASLLEDLSHIGLSQAPQLADEBPGRSIS-----GAAPSLGLDL	157
Db	806 SLGCEARS-SEDEDTATSGVTFLDSSDGPHTEKSIVALKSLQKVSTPPSLSLDI	864
Oy	158 LGPAT--GC-----LLDDGL-----EGLEPIDTSMYDELMAFAS	191
Db	865 PSSASDGCCEVLSPSAGPPGGQPRAVDSGVDTENYESPEFYLKAEHSESPEAFEPAS	924
Oy	192 EGLKPGEED-----GCKEAPPELDA---ELDYLMQVIYG-----TQALER	230
Db	925 EGESPGEPLLVSLGSLSKRSPRYDSAYPSDDLAESEPFGKEKHGIODSQEDLR	984
Oy	231 PPGPG 235	
Db	985 PPSPG 989	
RESULT	8	
H84310	cobryric acid synthase [imported] - Halobacterium sp. NRC-1	
C:	Species: Halobacterium sp. NRC-1	
C:	Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001	
C:	Accession: H84310	
R:	Mg, W.V.; Kennedy, S.P.; Mahairs, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, R.	
R:	Lethuiser, B.; Keller, K.; Crut, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.C.; Ja Jung, K.H.; Alam, M.; Freltas, T.	
Proc. Natl. Acad. Sci. U.S.A.	97, 12176-12181, 2000	
A:	Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;	
A:	Title: Genome sequence of Halobacterium species NRC-1.	
A:	Reference number: A84160; MID:20504483; PMID:11016950	
A:	Accession: H84310	
A:	Status: preliminary	
A:	Molecule type: DNA	
A:	Residues: 1-512 <STO>	
A:	Cross-references: GB:AE004437; NID:g10581058; PIDN:AA619852.1; GSPDB:GN00138	
C:	Genetics:	
A:	Gene: cbip	
C:	Superfamily: probable cobryric acid synthase	
Query Match	8.8%; Score 106.5; DB 2; Length 512;	
Best Local Similarity	24.5%; Pred. No. 3;	
Matches	70; Conservative 35; Mismatches 84; Indels 97; Gaps 17;	
Oy	27 DCGHAAVAQAAPPVAASSLFDSLVLKLHSLGQSEPDRLHLVLVNTLRRIQ--ASMAPA	84
Db	244 DPG-----LPADSVSILDPDPE-RVVDGQGDCVDPDAASVTAIVPHLPHTNSFTDLAPL	295
Oy	85 AALP-----PVPSP-----PAAPSVANILLASSDAL-	111
Db	296 AKTPGVAVAYQPLDAPLAADAVALPETKTFVTDLRAARAGLGRALRAFEDGEIVGLCGG	355
Oy	112 -----SASMASLLEDLSHIEGLSQAP--OP-----LADEGPGRSIGG	147
Db	356 YQMIGDRITDAARESTADDLGAVDGVGLEVEYFAFPDKRVETRELADCG--ALGG	411
Oy	148 AAPSGLALDI-----LGPAT--GCLLDGLGLGF-EDIDTSMTDNELMAP	189

DB 412 ATGANTGTEIMHGRRTTVDDGVPQPGPAAARGRLVGLTGLHGFENDAAKRGFRDVAFA 471
 QY 190 ASEGK-KPGPEDGPGKEAPELDEALDYLDVLTGTOALERPGR 234
 DB 472 A--GVEGPAPADTP--DQSPS-DAAAA--LVDAHVDLPLGVPPAP 510

RESULT 9
 746614
 chemotaxis protein cheA [imported] - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 31-Dec-2000
 C:Accession: T46614; F83463
 R:Kato, J.; Nakamura, T.; Kuroda, A.; Ohtake, H.
 Submitted to the EMBL Data Library, April 1998
 A:Description: Cloning, sequence and characterization of chemotaxis genes in Pseudomonas
 A:Reference number: 223079
 C:Accession: T46614

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-753 <KAT>
 A:Cross-references: EMBL:AB012767; PIDN:BAA33549.1
 A:Experimental source: strain PA01
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bz
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lim,
 ; Lorry, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: F83463
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-753 <STO>
 A:Cross-references: GB:AE004575; GB:AE004091; NID:99947404; PIDN:AAG04847.1; GSPDB:GNO01
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: cheA; PA1458
 C:Superfamily: chemotaxis protein cheA

Query Match 8.8%; Score 106.5; DB 2; Length 753;
 Best Local Similarity 25.7%; Pred. No. 4.7;
 Matches 59; Conservative 29; Mismatches 101; Indels 41; Gaps 10;

QY 2 LSKGKRRKEE--EKEPELVDSWMLDPGH-AAVAQAPPAVASSLTLSTLKHHSLO 58
 DB 119 LLAALARIAPPEGAPAPVQAPPAVPPAPPAEQASSDITDDEFEOQLDALQ 178
 DB 59 QSEPDRLHVLVVTLRITQASMAPAALPPVSPAPPAVADNLLASSDALSSMASL 118
 DB 179 GDE-----APASAVAPAPAPADDEISD-----AEFAL 207
 QY 119 LEDLSHIEGLSQAPQPLADEGP-PGRSIGAAPSLGALDILGPATCLLD--GLEGLFE 175
 DB 208 LDQL-HGKGFVPPAPVSAEPQAVPAVEPAALAAAGD-DISDEFALLDELHG-KGKFG 264
 QY 176 DIDSMTYNEMLMAPSEGLKRGPE-DGPGKEAPELDEALDYLDVLTG 224
 DB 265 DVPEAGTPAAPAAAAAPAAPEQKAPAAAGGDEISDDEFESLDELHG 314

RESULT 10
 S32538
 CGMP-gated cation channel 2, rod - human
 C:Species: Homo sapiens (man)
 C:Date: 22-Nov-1993 #sequence_revision 17-Nov-1995 #text_change 16-Jul-1999
 C:Accession: S32538
 R:Chan, T.Y.; Peng, Y.W.; Dallan, R.S.; Ahmed, B.; Reed, R.R.; Yau, K.W.
 Nature 362, 764-767, 1993
 A:Title: A new subunit of the cyclic nucleotide-gated cation channel in retinal rods.
 A:Reference number: S32538; MUID:93226050; PMID:7682292
 A:Accession: S32538
 A:Status: preliminary
 A:Molecule type: mRNA

A:Residues: 1-909 <CHE>
 C:Superfamily: CAMP receptor protein cyclic nucleotide-binding domain homology
 C:Keywords: CGMP binding
 F:620-742/Domain: CAMP receptor protein cyclic nucleotide-binding domain homology <CA

Query Match 8.8%; Score 106.5; DB 2; Length 909;
 Best Local Similarity 21.7%; Pred. No. 5.9;
 Matches 51; Conservative 34; Mismatches 89; Indels 61; Gaps 10;

QY 11 EEEKEPELVDSWMLDPGHAAVAQAPPAVASSLTLSTLKHHSLOQSEPDRLHVLV 70
 DB 88 EEEAKEP---QDW-----AETKEPEALAAASSGVPARKQHEVVEDT----- 131
 QY 71 VNTLRITQASMAPAALPPVSP--PPAPVADNLLASSDALSS--ASMASLDELISHI 125
 DB 132 -----ADSCPLAAEENPSTVLPSPSPAKSDTLIYSSASGTHRRKRLPSEDEAEL 183
 QY 126 EGLSQAPQPLADEGP-----GRTSGAAPSLGALDILGPATCLLDLEGF----- 174
 DB 184 KALSPAESPVVAWSDPTTPKDTDGODRAASTS-----TNSAITINDRLQELVLTFK 234
 QY 175 --EDIDTSMYDNEIM-----APASEGLKPGPEDGPGKEAPELDEALDYLT 218
 DB 235 ERTEVKRKLLIDPVTSDSESPKSPAKKAPAPADTPAPAE-PVEEHYCDML 288

RESULT 11
 A43776
 drebrin B2 - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 01-Dec-1993 #sequence_revision 30-Jan-1993 #text_change 01-Dec-2000
 C:Accession: A43776; B43776; I51212; I50221
 R:Kojima, N.; Kato, Y.; Shitao, T.; Oba, K.
 Brain Res. Mol. Brain Res. 4, 207-215, 1988
 A:Title: Nucleotide sequences of two embryonic drebrins, developmentally regulated br
 A:Reference number: A43776
 A:Accession: A43776
 A:Molecule type: mRNA
 A:Residues: 1-607 <KOU>
 A:Cross-references: GB:M36961; NID:9211725; PID:9211726
 A:Accession: B43776
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-315,359-607 <KO2>
 R:Kojima, N.; Shitao, T.; Oba, K.
 Brain Res. Mol. Brain Res. 19, 101-114, 1993
 A:Title: Molecular cloning of a developmentally regulated brain protein, chicken dreb
 A:Reference number: I51212; MUID:93368392; PMID:8361332
 A:Accession: I51212
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-132 <KO3>
 A:Cross-references: GB:S65230; NID:9410591; PID:AAB28010.1; PID:9410592

Query Match 8.7%; Score 105; DB 2; Length 607;
 Best Local Similarity 24.7%; Pred. No. 4.7;
 Matches 62; Conservative 18; Mismatches 77; Indels 94; Gaps 13;

QY 5 GLKRRKEE-----KEPLAVDSWMLDPGHAAVAQAPPAVASSLTLFDS 49
 DB 342 GEQRTAPAEETPATPKNSPSPQVAEPATQEHWPPEGEDKAAREP----- 391
 QY 50 VLKHHSLQOSEPD-----LRHLVLVVTLRITQASMAPAALPPVSPAPAS 98
 DB 392 -----DEDDPPRPAPWAGAVLGLV-----TLPESESPAPAPAS---EPQVETPG 436
 QY 99 VADNL--LASSDALSSMASLDELISHIEGLSQAPQPLADEGPGRSIGGAASGLALD 156
 DB 437 VAEPLIEIWMQSDGAAPAA-----TSTWLPDTPAGP-----VPEEGT-- 475
 QY 157 LIG-----PATGLDLDGLGLEFEDIDTSMYNEMLMA-----PASGLKPGPEDGPK 204
 DB 476 LIGDELPEPPATCDMAEQ-----HEVEEEEEEEETATACERHPTGLGYEGYQEGP-- 528

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1..901 <ZBO>
A:Cross-references: GI:863985; NID:q193208; PIDN:AAA37587.1; PID:q193210
A:Experimental source: Brain
A:Note: this sequence is inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:106578, NCBIPI:106579)
Laffer, E.; Zhou, S.; Sousa, R.; Tanner, N.H.

Qy 123 SHIEGSOAPPLAD-EGPPGRSIGAPSLGAL--DLGPGATGCLDDGEGLFEDIDT 179
| : | | : ||| : | | | | : : | | | : ||| : |

Db 133 GDEGSLQERQELSSGEGPGEAAAGLGLPSEGLASGEMQGAGGKVPKEAGVLGDSFV 19

OY 180 SMYDNELMAPASEGLKPGPED-----GPGK-EEAPELDEAELDYLMVYVGTQALERP 231
 Db 193 QGAAAEFTAPPEASGIPSSSEDEQIHTEEGEGKSGSPGPDHGPTELDGPD---GASAGEEP 249

RESULT 15

EDBE22

Immediate-early protein IER2.9 - bovine herpesvirus 1 (strain K22)

N:Alternate names: early protein ER2.6; p135 protein

C:Species: bovine herpesvirus 1

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 17-Mar-2000

C:Accession: A38209

R:Wirth, U.V.; Fraefel, C.; Vogt, B.; Vlcek, C.; Paces, V.; Schwyzler, M.

J. Virol. 66, 2763-2772, 1992

A:Title: Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1 are 3' coterm

A:Reference number: A38209; MUID:92219360; PMID:1313901

A:Accession: A38209

Molecule type: DNA

Accessions: 1-676 <MIR>

Cross-references: GB:M84464; NID:9330767; PIDN:AAA6061.1; PID:9330768

C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology

C:Keywords: DNA binding; immediate-early protein; transcription regulation; zinc finger

F:9-57/Domain: RING finger homology <RING>

F:13-51/Region: zinc finger C3HC4 motif

F:284-331/Region: acidic

Query Match 8.4%; Score 101.5; DB 1; Length 676;

Best Local Similarity 25.2%; Pred. No. 9.4;

Matches 63; Conservative 26; Mismatches 86; Indels 75; Gaps 11;

OY 7 KKKREEEKEEPLAVDSWMLDQGHAAVQAAPAVASSLSFDLSVLRHSLQSEFDLRH 66
 Db 462 RRRRTTELDRAPT-----PAPAPAPAPISTWIDLT----- 493
 OY 67 LVLVNTLRRIQASMAPAALP-----VPSPAPASVADNLLASSDAL 111
 Db 494 ----ANAPAR-PADPAPAAAGPALAGAOIGTPAAAAAVTAAAAAPSVARG--SAPSPAV 546
 OY 112 SASMASLLEDSHIGLSQAPQPLADEGPPGRSIGGAAPSIGALDLPATGCLDDGLE 171
 Db 547 TAAATGTAAAIIS----TRAPTP---SPAGRAPAADPRRAGAPALAGAAARAAGRGNP 597
 OY 172 GLFEDIDTSMYDNELMAPASEGLKPGPEDGPGKEAPELDEAELDYLMVYVGT-----Q 226
 Db 598 GERRRPASAM-----ARGDLDPGPSSSAOKRRRTTEAEVAA--WRESLIGTPRRSSA 647
 OY 227 ALERPPGPGR 236
 Db 648 ALAPQPG-GR 656

Search completed: August 19, 2003, 09:32:51

Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 85 Seconds

(without alignments)
440.700 Million cell updates/sec

Title: US-09-701-675A-3

Perfect score: 1210

Sequence: 1 MTSKLRKRREREKEPLA.....YLMYLVGTQLERPPGGR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: A_Geneseq_19Jun03:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1210	100.0	236	21 AAY44362 Human cell cycle r
2	1210	100.0	278	22 AAM25550 Human protein sequ
3	1206	99.7	236	22 AAM93724 Human polypeptide,
4	1206	99.7	236	22 AAG89292 Human secreted pro
5	1203	99.4	236	20 AAY31829 Human adult blood
6	1198	99.0	236	20 AAY02619 Amino acid sequenc
7	1164	96.2	236	20 AAY36004 Extended human sec
8	335.5	27.7	116	22 ABG14355 Novel human diagno
9	289	23.9	58	20 AAY12453 Human 5' EST secre

10	213	17.6	196	22 AAM93922 Human polypeptide,
11	213	17.6	196	22 AAB35402 Replication protei
12	208.5	17.2	314	22 AAB65076 Hypoxia-induced pr
13	170	14.0	241	22 AAM39725 Human polypeptide
14	170	14.0	254	22 AAM41511 Human polypeptide
15	168	13.9	254	21 AAB58258 Lung cancer associ
16	168	13.9	254	22 AAG73682 Human colon cancer
17	165	13.6	241	22 AAO2898 Human haematopoiet
18	165	13.6	241	22 AAB06375 Human COL protein
19	163.5	13.5	157	22 AAG74742 Human chd protein
20	152	12.6	237	23 AAO2897 Mouse haematopoiet
21	129.5	10.7	322	22 ABB22868 Novel human diagno
22	116	9.6	644	22 ABB83698 Human prostate can
23	116	9.6	789	22 AAB83690 Human prostate can
24	115	9.5	255	22 AAU18154 Human DNA-binding
25	115	9.5	255	23 ABB92575 Human DNA-binding
26	114	9.4	42	23 ABB69181 Human polypeptide
27	114	9.4	2038	23 AAB25098 Human kinase and p
28	114	9.4	2161	22 AAM78959 Human protein SEQ
29	114	9.4	2189	22 AAM79943 Human protein SEQ
30	114	9.4	2429	22 AAE21713 Human PKIN-8 prote
31	114	9.4	2523	22 AAU03503 Human protein kina
32	114	9.4	2545	23 ABB98406 Human NOV, MAST70
33	113	9.3	983	20 AAY09513 Mouse JMT protein
34	112	9.3	791	23 ABB65089 Hypoxia-repressed
35	112	9.3	802	22 ABB28366 Novel human diagno
36	108.5	9.0	248	22 ABB04071 Novel human diagno
37	108.5	9.0	1101	22 AAB04071 Novel GTPase activ
38	108.5	9.0	1219	22 AAM50136 Novel human diagno
39	108.5	9.0	1279	22 AAM78969 Human protein SEQ
40	108.5	9.0	1464	22 AAM41039 Human polypeptide
41	107.5	8.9	1444	22 ABB15667 Novel human diagno
42	106.5	8.8	663	18 AAM29150 Dual-specific murt
43	105	8.7	175	22 ABB15339 Novel human diagno
44	104	8.6	113	23 AAO22900 N-terminal region
45	104	8.6	443	23 ABB40426 Human secreted pro

ALIGNMENTS

RESULT 1	AY44362	AY44362 standard; protein; 236 AA.
ID	AY44362	
XX	AY44362	
AC	AY44362	
XX	14-MAR-2000 (first entry)	
DT	14-MAR-2000 (first entry)	
XX		
DE	Human cell cycle regulation protein-3.	
XX		
KW	CECRP-3; cell cycle regulation protein-3; cell proliferation;	
KW	cell proliferative disease; cancer; atherosclerosis; cirrhosis;	
KW	hepatitis; psoriasis; immune system disorder; allergy; asthma;	
KW	acquired immune deficiency syndrome; Crohn's disease; Blast method;	
KW	rheumatoid arthritis; gene therapy; chromosomal mapping.	
XX		
OS	Homo sapiens.	
XX		
FX		
FT	Key	Location/Qualifiers
FT	Modified-site	44 /note= "Potential phosphorylation site"
FT	Modified-site	60 /note= "Potential phosphorylation site"
FT	Modified-site	73 /note= "Potential phosphorylation site"
FT	Modified-site	98 /note= "Potential phosphorylation site"
FT	Modified-site	117 /note= "Potential phosphorylation site"
FT	Modified-site	123 /note= "Potential phosphorylation site"
FT	Modified-site	180 /note= "Potential phosphorylation site"

FT /note= "Potential phosphorylation site"
 XX
 XX W0964593-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99MO-US12906.
 XX
 PR 08-JUN-1999; 98US-0088695.
 XX
 PA (INCYTE) INCYTE PHARM INC.
 XX
 PI Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR,
 PI Patterson C;
 DR WPI; 2000-105887/09.
 DR N-PSDB; AA229482.
 XX
 PT Novel regulatory proteins, for diagnosis, treatment and prevention of
 PT cell proliferative and immune system diseases
 XX
 PS Claim 1; Page 70; 88pp; English.
 XX
 CC The present sequence is cell cycle regulation protein-3 (CECRP-3). Prints
 CC analytical method was used to identify this protein. CECRPs are
 CC activators of cell proliferation or inhibitors of cellular processes that
 CC modulate proliferation. They are used to treat or prevent cell
 CC proliferative diseases like cancers, atherosclerosis, cirrhosis,
 CC hepatitis, psoriasis, immune system disorders (e.g. acquired immune
 CC deficiency syndrome, allergy, asthma, Crohn's disease,
 CC rheumatoid arthritis). Antibodies are raised to screen for specific
 CC binding agents. The corresponding nucleic acid is used in gene therapy,
 CC chromosomal mapping and isolation of related sequences.
 CC
 SQ Sequence 236 AA;
 Query Match 100.0%; Score 1210; DB 21; Length 236;
 Best Local Similarity 100.0%; Pred. No. 2.4e-93;
 Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLSGKLRKRREERKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVLKLRHSIQOS 60
 DB 1 MLSGKLRKRREERKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVLKLRHSIQOS 60
 QY 61 EPPDRHLVLYVNTLRRIQASMAAPAAALPPVSPPAASVADNLLASSDAISASMASILE 120
 DB 61 EPPDRHLVLYVNTLRRIQASMAAPAAALPPVSPPAASVADNLLASSDAISASMASILE 120
 QY 121 DLSHTEGSQLAPPLADEGPPGRSTIGGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
 DB 121 DLSHTEGSQLAPPLADEGPPGRSTIGGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
 QY 181 MYDNELMAPASEGLKRPEDPGKGEAPDELDEALDYLMADVLYGQALERPGR 236
 DB 181 MYDNELMAPASEGLKRPEDPGKGEAPDELDEALDYLMADVLYGQALERPGR 236
 RESULT 2
 AAM25550
 ID AAM25550 standard; Protein: 278 AA.
 XX
 AC AAM25550;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:1065.
 XX
 KW Human: cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiallergic; immunosuppressive;
 KW antibacterial; endocrine; cardiac; central nervous system; vitruide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antineoplastic; anaemia;
 KW antiagregant; haemostatic; vulnary; antiallergic; osteopathic; eczema;
 KW dermatological; antiallergic; antiaesthetic; antidiabetic; cytostatic;

KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.
 XX
 OS Homo sapiens.
 XX
 XX W0200153455-A2.
 XX
 PD 26-JUL-2001.
 XX
 PF 22-DEC-2000; 2000MO-US35017.
 XX
 PR 23-DEC-1999; 99US-0471275.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 PI
 DR WPI; 2001-457603/49.
 DR N-PSDB; AAH99491.
 XX
 PT Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection
 XX
 PS Claim 20; Page 214; 1217pp; English.
 XX
 CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 CC AAM25963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
 CC central nervous system; vitruide; anti-HIV; fungicide; antineoplastic;
 CC cardiovascular; antineoplastic; antiagregant; haemostatic; vulnary;
 CC antiallergic; osteopathic; dermatological; antiallergic; antiaesthetic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 CC
 SQ Sequence 278 AA;
 Query Match 100.0%; Score 1210; DB 22; Length 278;
 Best Local Similarity 100.0%; Pred. No. 3e-93;
 Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLSGKLRKRREERKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVLKLRHSIQOS 60
 DB 43 MLSGKLRKRREERKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVLKLRHSIQOS 102
 QY 61 EPPDRHLVLYVNTLRRIQASMAAPAAALPPVSPPAASVADNLLASSDAISASMASILE 120
 DB 103 EPPDRHLVLYVNTLRRIQASMAAPAAALPPVSPPAASVADNLLASSDAISASMASILE 162
 QY 121 DLSHTEGSQLAPPLADEGPPGRSTIGGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
 DB 163 DLSHTEGSQLAPPLADEGPPGRSTIGGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 222

QY 181 MYDNEIMAPASEGLKPGEDGPKGEAPBELDEALDYIMDVLTGTOALERPGR 236
DB 223 MYDNEIMAPASEGLKPGEDGPKGEAPBELDEALDYIMDVLTGTOALERPGR 278

RESULT 3
AAM93724

ID AAM93724 standard; Protein: 236 AA.

AC AAM93724;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 3677.

KM Human; full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

PN EP130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

PA (HELI-) HELIX RES INST.

PI Ohta T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR N-PSDB; AAK94674.

PT 830. Primers useful for synthesizing full length cDNA clones and their

PS use in genetic manipulation -

CC Claim 8; SEQ ID NO 3677; 1380pp + sequence listing; English.

CC The invention relates to primers for synthesizing full length cDNA

CC clones. 830 cDNA molecules encoding a human protein have been

CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

CC molecules have been determined. Primers for synthesizing the full length

CC cDNA are useful for clarifying the function of the protein encoded by

CC the cDNA. The full length clones were obtained by construction of full

CC length enriched cDNA libraries that were synthesised by the oligo-capping

CC method. The primers enable the production of the full length cDNA easily

CC without any special methods. The present sequence is a polypeptide

CC encoded by a full length human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in CD-ROM format directly from EPO.

SQ Sequence 236 AA;

QY Query Match 99.7%; Score 1206; DB 22; Length 236;

DB Best Local Similarity 99.6%; Pred. No. 5.3e-93;

Matches 235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 181 MYDNEIMAPASEGLKPGEDGPKGEAPBELDEALDYIMDVLTGTOALERPGR 236
DB 181 MYDNEIMAPASEGLKPGEDGPKGEAPBELDEALDYIMDVLTGTOALERPGR 236

RESULT 4
AAG89292

ID AAG89292 standard; Protein: 236 AA.

AC AAG89292;

DT 11-SEP-2001 (first entry)

DE Human secreted protein, SEQ ID NO: 412.

KM Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;

OS GENSET.

PN MO200142451-A2.

PD 14-JUN-2001.

PF 07-DEC-2000; 2000MO-IB01938.

PR 08-DEC-1999; 99US-0169629.

PR 06-MAR-2000; 2000US-0187470.

PA (GENSET) GENSET.

PI Dumas Milne Edwards J, Bougueleret L, Jobert S;

PI WPI: 2001-367870/38.

DR N-PSDB; AAK64895.

PT Full length GENSET human nucleic acids encoding potentially secreted

PS proteins, useful in gene therapy and vaccination against a variety of

CC diseases, and for diagnosis of those diseases -

CC Claim 21; Page 889-890; 921pp; English.

CC The invention relates to full length GENSET human nucleic acids encoding

CC potentially secreted proteins. The nucleic acids and the polypeptides

CC they encode may be used in the prevention, treatment and diagnosis of

CC diseases associated with inappropriate GENSET gene expression. For

CC example, they be used to treat disorders associated with decreased

CC GENSET gene expression by rectifying mutations or deletions in a

CC patient's genome that affect the activity of GENSET or by supplementing

CC the patient's own production of GENSET polypeptides. Conversely,

CC antisense nucleic acid molecules may be administered to down regulate

CC GENSET expression by binding with the cells' own genes and preventing

CC their expression. The sense and antisense nucleic acids may also be

CC used as DNA probes in diagnostic assays to detect and quantitate the

CC presence of similar nucleic acid sequences in samples, and hence to

CC determine which patients may be in need of restorative therapy.

CC The GENSET polypeptides may be used as antigens in the production of

CC antibodies and in assays to identify modulators (agonists and

CC antagonists) of GENSET polypeptide expression and activity. The

CC present sequence is a GENSET polypeptide of the invention.

SQ Sequence 236 AA;

QY Query Match 99.7%; Score 1206; DB 22; Length 236;

DB Best Local Similarity 99.6%; Pred. No. 5.3e-93;

Matches 235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

DB 61 EEDLRLHLYVNTLRRIQASMAPPAALPPVSPPAAPSVADNLASSDAALASMASLLE 120
QY 121 DLSHIEGLSQAPQPLADEGPPGRSIGAAPSLGALDLGPATGCLLDGLEGLEFEDIDTS 180
DB 121 DLSHIEGLSQAPQPLADEGPPGRSIGAAPSLGALDLGPATGCLLDGLEGLEFEDIDTS 180
QY 181 MYDNLMPASBGLKPGPEDGPKKEAPLDEALDYLMDVLYGTQALERPPGGR 236
DB 181 MYDNLMPASBGLKPGPEDGPKKEAPLDEALDYLMDVLYGTQALERPPGGR 236

RESULT 5
AA31829
ID AAY31829 standard; Protein; 236 AA.
XX
AC AAY31829;
XX
XX 06-DEC-1999 (first entry)
DT
XX
XX Human adult blood secreted protein g21_1.
DE
XX Secreted protein; g21_1; human; therapy; diagnosis; vaccine; blood.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX Peptide 19..21 /note= "signal peptide"
XX Peptide 67..79 /note= "alternative signal peptide"
XX Protein 32..236 /note= "mature protein"
XX Protein 80..236 /note= "alternative mature protein"
XX Misc-difference 1 /note= "encoded by AWG"
XX Misc-difference 137 /note= "encoded by GAS"
XX Domain 40 /note= "a putative transmembrane domain is centered around this residue"
XX Domain 80 /note= "a putative transmembrane domain is centered around this residue"
XX Domain 150 /note= "a putative transmembrane domain is centered around this residue"
XX Domain /note= "a putative transmembrane domain is centered around this residue"
XX W09947555-A1.
XX
XX 23-SEP-1999.
PD
XX
XX 18-MAR-1999; 99WO-US05939.
XX
XX 20-MAR-1998; 98US-0078803.
XX
XX 17-MAR-1999; 99US-0078803.
XX
XX (GENY ) GENETICS INST INC.
XX
XX Jacobs R, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Steindler RJ;
XX WPI; 1999-562059/47.
XX N-PSDB; AA219894.
DR
XX
XX New polynucleotides derived from murine foetal cell cDNA libraries,
XX potentially used as, e.g. vaccines -
XX
XX Claim 13(a); Page 94; 107pp; English.
XX
XX This is the predicted amino acid sequence of a novel human secreted
XX protein, g21_1, as deduced from an isolated adult blood cDNA
XX

```

```

CC clone (see AA219894). The invention provides new human secreted
CC proteins (see AAY31828-38) and polynucleotides (see AA219893-901)
CC isolated from foetal cell, adult blood, adult brain and foetal
CC kidney cDNA libraries. They are predicted to have biological
CC activities which would make them suitable for treating, preventing
CC or ameliorating medical conditions in humans and animals, although
CC no supporting data are given. Suggested activities include
CC nutritional, cytokine, tissue growth, cell proliferation and
CC differentiation, immunostimulant (e.g. as vaccine),
CC immunosuppressive, haematopoiesis regulating, activin or inhibin,
CC chemotactic or chemokinetic, haemostatic or thrombolytic,
CC receptor/ligand activity, antiinflammatory, catheterin or tumour
CC invasion suppressor, and tumour inhibition activities.
XX
XX Sequence 236 AA;
SQ

Query Match 99.4%; Score 1203; DB 20; Length 236;
Best Local Similarity 99.6%; Pred. No. 9.4e-93;
Matches 235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVTLKHSIQDS 60
DB 1 MLSKGLKRRKEEKEEKEEPLAVDSWMLDPGHAAVAQAPPAVASSSLFDLSVTLKHSIQDS 60
QY 61 EEDLRLHLYVNTLRRIQASMAPPAALPPVSPPAAPSVADNLASSDAALASMASLLE 120
DB 61 EEDLRLHLYVNTLRRIQASMAPPAALPPVSPPAAPSVADNLASSDAALASMASLLE 120
QY 121 DLSHIEGLSQAPQPLADEGPPGRSIGAAPSLGALDLGPATGCLLDGLEGLEFEDIDTS 180
DB 121 DLSHIEGLSQAPQPLADEGPPGRSIGAAPSLGALDLGPATGCLLDGLEGLEFEDIDTS 180
QY 181 MYDNLMPASBGLKPGPEDGPKKEAPLDEALDYLMDVLYGTQALERPPGGR 236
DB 181 MYDNLMPASBGLKPGPEDGPKKEAPLDEALDYLMDVLYGTQALERPPGGR 236

RESULT 6
AA302619
ID AAY02619 standard; Protein; 236 AA.
XX
XX AAY02619;
XX
XX 21-JUL-1999 (first entry)
DT
XX
XX Amino acid sequence of a p16 binding protein.
XX
XX DE
XX
XX p16BP1; cell cycle regulator protein p16; binding protein;
XX cell cycle regulation; cyclin D dependent kinase; CDK4;
XX cell cycle progression; 19q13.1-13.2; cancer; gene therapy;
XX cell cycle disorder.
XX
XX Homo sapiens.
OS
XX
XX W09923218-A1.
XX
XX 14-MAY-1999.
PD
XX
XX 04-NOV-1998; 98WO-JP04972.
XX
XX 05-NOV-1997; 97JP-0302564.
XX
XX (SOME ) SUMITOMO ELECTRIC IND CO.
XX
XX Hara E, Nakamura T;
XX WPI; 1999-313340/26.
XX N-PSDB; AAX36177.
DR
XX
XX New p16-binding protein useful in gene therapy of cancer and other
XX cell cycle disorders
XX
XX Claim 1; Page 34-35; 40pp; Japanese.
PS

```

XX The present sequence represents a protein, designated p16p1, which binds
 CC to the cell cycle regulator protein p16. p16p1 plays a role in cell
 CC cycle regulation by binding to p16 which is a protein binding to a
 CC cyclin D dependent kinase (CDK4) involved in the regulation of cell
 CC cycle progression. The gene encoding p16p1 has been localized to
 CC 19q13.1-13.2. The binding protein, polynucleotides encoding it, and
 CC antibodies recognizing it are useful in the investigation of cell cycle
 CC progression and regulation, particularly in connection with cancer
 CC formation, for gene therapy of cancer and other cell cycle disorders,
 CC and for diagnosis of such disorders. The sense or antisense
 CC polynucleotides are useful as probes to screen a cDNA library.

SQ Sequence 236 AA:

Query Match 99.0%; Score 1198; DB 20; Length 236;
 Best Local Similarity 99.2%; Pred. No. 2,5e-92;
 Matches 234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLSKGLKRRREEEKEPLAVDSWMLDPGHAAVAQAAPVAVASSSLFDLSVTLKHSLSQGS 60
 DB 1 MLSKGLKRRREEEKEPLAVDSWMLDPGHAAVAQAAPVAVASSSLFDLSVTLKHSLSQGS 60
 QY 61 EPDLRLVLYVNTLRRIQASMAPAALPPVPSPPAPSVADNMLASSDAALSASMASLLE 120
 DB 61 EPDLRLVLYVNTLRRIQASMAPAALPPVPSPPAPSVADNMLASSDAALSASMASLLE 120
 QY 121 DLSHIEGLSQAPQPLADEGPPGSRISGAPSLGALDILGPATGCLLDLDEGLFEDIDTS 180
 DB 121 DLSHIEGLSQAPQPLADEGPPGSRISGAPSLGALDILGPATGCLLDLDEGLFEDIDTS 180
 QY 181 MYDNELMAPASGLKPGPEDGPGKEAPELDEAELDYLMVLYGTQALERRPPGGR 236
 DB 181 MYDNELMAPASGLKPGPEDGPGKEAPELDEAELDYLMVLYGTQALERRPPGGR 236

RESULT 7
 AA36004
 ID AAY36004 standard; Protein; 236 AA.

AC AAY36004;
 XX
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Extended human secreted protein sequence, SBO ID NO. 389.
 XX
 KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
 KW cellular differentiation; immune system regulator; anti-inflammatory;
 KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
 KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
 KW genetic disease.
 XX
 OS Homo sapiens.
 XX
 PN WO9931236-A2.
 PD 24-JUN-1999.
 XX
 PF 17-DEC-1998; 98WO-1B02122.
 XX
 PR 10-AUG-1998; 98US-0096116.
 PR 17-DEC-1997; 97US-0069957.
 PR 09-FEB-1998; 98US-0074121.
 PR 13-APR-1998; 98US-0081563.
 XX
 PA (GSEF) GENSEF.
 XX
 PI Bougueleret L, Duclet A, Dumas Mline Edwards J;
 XX
 DR WPI; 1999-385906/32.
 DR N-PSDB; AAX97688.
 XX
 PT New isolated human secreted proteins

XX Claim 9; Page 330-331; 516pp; English.
 PS
 XX This sequence is encoded by an extended human secreted protein coding
 CC sequence of the invention. The secreted proteins can be used in treating
 CC or controlling a variety of human conditions. The secreted proteins may
 CC act as cytokines or may affect cellular proliferation or differentiation
 CC or may act as immune system regulators, haematopoiesis regulators, tissue
 CC growth regulators, regulators of reproductive hormones or cell movement
 CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
 CC tumour inhibition activity. The DNAs can be used in forensic procedures
 CC to identify individuals or in diagnostic procedures to identify
 CC the genes corresponding to the extended cDNAs. They are also useful for
 CC constructing a high resolution map of the human chromosomes. They can
 CC also be used for gene therapy to control or treat genetic diseases.

SQ Sequence 236 AA:

Query Match 96.2%; Score 1164; DB 20; Length 236;
 Best Local Similarity 95.8%; Pred. No. 1.7e-89;
 Matches 226; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 MLSKGLKRRREEEKEPLAVDSWMLDPGHAAVAQAAPVAVASSSLFDLSVTLKHSLSQGS 60
 DB 1 MLSKGLKRRREEEKEPLAVDSWMLDPGHAAVAQAAPVAVASSSLFDLSVTLKHSLSQGS 60
 QY 61 EPDLRLVLYVNTLRRIQASMAPAALPPVPSPPAPSVADNMLASSDAALSASMASLLE 120
 DB 61 EPDLRLVLYVNTLRRIQASMAPAALPPVPSPPAPSVADNMLASSDAALSASMASLLE 120
 QY 121 DLSHIEGLSQAPQPLADEGPPGSRISGAPSLGALDILGPATGCLLDLDEGLFEDIDTS 180
 DB 121 DLSHIEGLSQAPQPLADEGPPGSRISGAPSLGALDILGPATGCLLDLDEGLFEDIDTS 180
 QY 181 MYDNELMAPASGLKPGPEDGPGKEAPELDEAELDYLMVLYGTQALERRPPGGR 236
 DB 181 MYDNELMAPASGLKPGPEDGPGKEAPELDEAELDYLMVLYGTQALERRPPGGR 236

RESULT 8
 ABG14355
 ID ABG14355 standard; Protein; 116 AA.

AC ABG14355;
 XX
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #14346.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS78542.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID NO 44714; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIRO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 116 AA;

Query Match 27.7%; Score 335.5; DB 22; Length 116;
Best Local Similarity 69.5%; Pred. No. 1.6e-20;
Matches 73; Conservative 5; Mismatches 14; Indels 13; Gaps 4;

OY 139 GPGRSIGGAAPSLGALDLGPAAGCLLDGSLGFEFDIDTSMYDNE-----LWAPASEG 193
DB 3 GPOASSIGGAAPSGALDLDLPATGCLLDGSLGFEFDIDTSMYDNEELGTSWMAASQA 62
OY 194 LKPGEDGPG-KKEAPFELDEAELDYLMVYVGTQAL---EPPPG 234
DB 63 LK---DGPQGRKRLKSLDEAELDYLMVYVGTQATGSDRGPRP 103

RESULT 9
AA12453
ID AA12453 standard; Protein: 58 AA.
XX
AC AA12453;
XX
DT 17-JUN-1999 (first entry)
XX
DE Human 5' EST secreted protein SEQ ID NO:484.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX forensic; gene therapy; chromosome mapping; signal peptide;
XX upstream regulatory sequence; cytokine activity; cell proliferation;
XX differentiation; haematopoiesis regulation; tissue growth regulation;
XX reproductive hormone regulation; chemokine; chemokine; haemostatic;
XX thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS Homo sapiens.
XX
PN MO906548-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01222.
XX
PR 01-AUG-1997; 97US-0905135.
XX
PA (BEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX

DR WPI: 1999-153778/13.
DR N-PSDB: AAX41286.
XX
XX

PT New nucleic acids encoding human secreted proteins - obtained from
PT CDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
XX
XX
PS Claim 27; Page 782-783; 824pp; English.

XX AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
XX human secreted proteins, and encode the proteins given in AA12561 to
XX AA12564, respectively. The proteins given represent the signal peptide
XX and an N-terminal fragment of a secreted protein. The nucleic acid
XX sequences can be used for producing secreted human gene products. They
XX can also be used to develop products for diagnosis and therapy. The
XX proteins obtained may have cytokine activity, cell
XX proliferation/differentiation activity, haematopoiesis regulating
XX activity, tissue growth regulating activity, reproductive hormone
XX regulating activity, chemokine/chemokine activity, haemostatic and
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, tumour inhibition activity or other activities. The products
XX can be used in forensic, gene therapy and chromosome mapping procedures.
XX The sequences can also be used for obtaining corresponding promoter
XX sequences. The nucleic acids encoding the signal peptide can be used for
XX directing extracellular secretion of a polypeptide or the insertion of a
XX polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 58 AA;

Query Match 23.9%; Score 289; DB 20; Length 58;
Best Local Similarity 100.0%; Pred. No. 5.4e-17;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MLSKGLKRREREKEPLAVDSWMLDPGHAAPPAVASSSLFDLSYLKTHS 56
DB 1 MLSKGLKRREREKEPLAVDSWMLDPGHAAPPAVASSSLFDLSYLKTHS 56

RESULT 10
AAM93922
ID AAM93922 standard; Protein: 196 AA.
XX
AC AAM93922;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 4085.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
XX
PR 11-JAN-2000; 2000JP-0118774.
XX
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
XX
DR N-PSDB: AAK94884.
XX

PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX

CC The present invention relates to methods for identifying genes and
CC proteins that are implicated in a specific disease or physiological
CC condition. The method comprises comparing the transcriptome/proteome of a
CC specialised cell type implicated in a disease or condition with that of a
CC second specialised cell type, under two experimental conditions, and
CC identifying a gene that is differentially regulated in the two
CC specialised cell types under experimental conditions. ABV77873-ABV78116
CC and ABP65061-ABP65257 were identified using the methods of the invention.
CC The coding sequences and proteins are useful for treating a disease in a
CC patient, for manufacture of a medicament for treating hypoxia-regulated
CC conditions, and for regulating tumourigenesis, angiogenesis, apoptosis,
CC biological response to hypoxia conditions, or hypoxic-associated
CC pathology in a patient. The coding sequences and proteins are also useful
CC for monitoring the therapeutic treatment of a disease or physiological
CC condition, such as cancer, ischaemic conditions, reperfusion injury,
CC retinopathy, neonatal stress, pre-eclampsia, atherosclerosis, inflammatory
CC conditions, wound healing, inflammation, erythropoiesis or hair loss.
CC
CC xx

Sequence	314 AA;	17.2%;	Score 208.5;	DB 23;	Length 314;
Query Match		27.2%;	Fred. No. 2.3e-09;		
Best Local Similarity		37;	Mismatches 91;		
Matches	86;	Conservative	91;	Indels 107;	Gaps 11

```

QY      1  MLSGKRRKEEKEEPLAVDSWMLDGGHAAVAQAPAVA-----SSSLFDLSYVLKHS 56
Db      1  MLGKGKKRKEEHEH-----GLEKIVSPCDGPKSVSTLDRQTIFNLSKLKYNH 51

QY      57  LQOSEPDLRLVLYVNTLRIRIQASKAPAAALPYVSPSPAASV-----ADNLTA 105
Db      52  RPLTEPSLOKTVLLNNMLRIQOELKQSGSLRPMETPSQPTTESDSYREAPAFASHIA 111

QY      106  S-----SDAALSAG--ASLLEDSHTEGLSQAOBPLADE--GPPG-----142
Db      112  SPSSHPCDLGSTTPEACLTFAASLLEDDDDTFCISQAQOPAPATLSPALLPENKSS 171

QY      143  -----RSIGCAPSLGALDLP-----160
Db      172  ALDEIEELCPSTSTEATAATDSYKGSSSAGCTOKLDGPPESRADSKLMDSLPGNFEI 231

QY      161  --ATGCLDDDLLEG-LFEDIDTSMYNDNLMAAPAS---EGLKGPEDGEGKEAP-----208
Db      232  TTSTGFLDULDLIDILFADIDTSMYNDPFCPCISSSOTASKMAPVSADDLKTLAPISSQPV 291

QY      209  -----ELDEAEULYMDLYVT 225
Db      292  TPQGPFRKDLTELPHIMEVLVGS 314

```

	RESULT	13
	AAM39725	
ID	AAM39725	standard; Protein: 241 AA.
XX		
AC	AAM39725;	
XX		
DT	22-OCT-2001	(first entry)
XX		
DE	Human polypeptide SEQ	ID NO 2870.
XX		
KW	Human; neotropic; immunosuppressant; cyostatic; gene therapy; cancer;	
KW	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	amorphotic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	leukaemia.	
XX		
OS	Homo sapiens.	
XX		
PN	MO20015312-A1.	
XX		
PD	26-JUL-2001.	
XX		
FE	26-DEC-2000; 2000WO-US34263.	

XX	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0558042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	
DR	WPI: 2001-442253/47.
DR	N-PSDB; AAI58881.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	
XX	Example 4; SEQ ID NO 2870; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAM38642-AAM42213) with neurotrophic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, CC Activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic CC and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukemias and CC C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification.

xx

Sequence	241 AA;
----------	---------

Query Match	14.0%	Score 170;	DB 22;	Length 241;
Best Local Similarity	28.0%;	Pred. No. 2.7e+06;		
Matches 74; Conservative	39;	Mismatches 89;	Indels 62;	Gaps 13;

[illegible]

XX 22-OCT-2001 (first entry)
 DT Human polypeptide SEQ ID NO 6442.
 DE
 XX
 KW Human: neurotrophic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW Leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 XX
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSRO INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AA160667.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2; SEQ ID NO 6442; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 254 AA;
 QY Query Match 14.0%; Score 170; DB 22; Length 254;
 QY Best Local Similarity 28.0%; Pred. No. 2.9e-06;
 Db Matches 74; Conservative 39; Mismatches 89; Indels 62; Gaps 13;
 QY 1 M L S K G I K R K - R E E E K E P L A - - - V D S W M I D P G A A V A Q A P P A V A S S I E D L S Y L K L H 54
 Db 14 M F A R G K R R C V G H E E D E G A L G L K T V S S Y - - - - - S I O R S I L D M S I V K I Q 59
 QY 55 H S I O S E P D L R H L V L V N T L R I O A S M A P A A L P V P S P P A P S V A N D M L A S S D A L S 114
 Db 60 L C H M L V E P M C R S V L I A N T V R I O E M T O D G T W R T V - A P O A A E R A P L D R L V S T E I L C R A A 118

QY 115 M - - - - - A S L E D L S H I R G - - - - - L S Q A P O P L A D E G - - - - - P P G R S I G G A P S I G - 153
 Db 119 W G E G A H P A S G L G D - G H T G G P V S D L C P V T S A Q A P R H L O S S A M E M D P R E N R G S F H N S L D Q 177
 QY 154 - - - A L L L G P A T C C L D D G L E G F E D I D T S M Y P - N E I M A P A S G L K P G P D G - - - - - 201
 Db 178 I F E T L E T K N S - C - - - - M E E L F S D V S P Y D L D V L G M G G A R G P G E G L A P A T 230
 QY 202 P G K E A P E L D E A L D Y I M D V L G T 225
 Db 231 P G P S S C K S D L G E L D H V E I L V E T 254
 RESULT 15
 AAB58258
 ID AAB58258 standard; Protein; 254 AA.
 XX
 AC AAB58258;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Lung cancer associated polypeptide sequence SEQ ID 596.
 XX
 KW Human: lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnary;
 KW gastrointestinal; nephrotropic; antineoplastic; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease.
 XX
 OS Homo sapiens.
 XX
 PN WO20005180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Ruben SM;
 XX
 DR WPI: 2000-587514/55.
 DR N-PSDB; AAF18134.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 11: Page 1089-1090; 1425bp; English.
 XX
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAF58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences; their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnary; gastrointestinal
 CC general; nephrotropic; antineoplastic; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, vulnary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 CC
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 11 Seconds

(without alignments)
1008.937 Million cell updates/sec

Title: US-09-701-675A-3

Perfect score: 1210
Sequence: 1 MLSKGLKKRKEEKEEPLA.....YIMDVLTGTLERPPGGR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	208.5	17.2	314	1 Y127_HUMAN	Q14140 homo sapien
2	120	9.9	3164	1 TEGU_HSV11	P10720 herpes simp
3	118.5	9.8	2161	1 SHK1_HUMAN	Q9Y566 homo sapien
4	112.5	9.3	652	1 DRB8_CHICK	P18302 gallus gall
5	112	9.3	863	1 M113_HUMAN	Q8N3F8 gallus gall
6	108.5	9.0	488	1 CTA0_HUMAN	Q96M44 homo sapien
7	107.5	8.9	440	1 FXGA_CHICK	Q98337 gallus gall
8	107.5	8.9	708	1 VP40_HCMVA	P16753 human cytom
9	106.5	8.8	663	1 DUS8_MOUSE	O09112 mus musculu
10	106.5	8.8	909	1 CNG4_HUMAN	Q14028 homo sapien
11	105	8.7	447	1 ODP2_RHIME	Q919N3 rhizobium m
12	103.5	8.6	1567	1 FMN2_MOUSE	Q91104 mus musculu
13	102.5	8.5	901	1 A180_MOUSE	O61548 mus musculu
14	102.5	8.5	908	1 SRC4_RABIT	P13666 oryctolagus
15	102	8.4	1516	1 CA1H_HUMAN	P39060 homo sapien
16	101.5	8.4	676	1 ICP0_HSVK	P29836 bovine hepp
17	101.5	8.4	2774	1 MAPA_RAT	P34926 ratius norv
18	101	8.3	433	1 AFLR_EMENT	P52957 emericella
19	101	8.3	1433	1 Y310_HUMAN	O15027 homo sapien
20	100.5	8.3	676	1 ICP0_HSVB1	P29128 bovine hepp
21	100	8.3	655	1 FXO1_HUMAN	Q12778 homo sapien
22	100	8.3	806	1 MK07_MOUSE	Q9WVS8 mus musculu
23	100	8.3	1175	1 HCN4_RABIT	Q9TV66 oryctolagus
24	100	8.3	1509	1 GSR1_HUMAN	O9NMZ4 homo sapien
25	99.5	8.2	359	1 CEB4_MOUSE	P53566 mus musculu
26	99.5	8.2	1106	1 IPD0_BOVIN	P28339 bos taurus
27	99	8.2	373	1 HMT1_DROST	Q24648 drosophila
28	99	8.2	498	1 VGLX_PRIV1	P07562 pseudorabla
29	98.5	8.1	281	1 PMXA_RAT	Q6W782 ratius norv
30	98.5	8.1	2167	1 SHK1_RAT	Q9W48 ratius norv
31	98	8.1	824	1 JIP2_HUMAN	O13387 homo sapien
32	98	8.1	1199	1 P121_RAT	P52591 ratius norv
33	98	8.1	3010	1 POLG_HCVTV	P29846 h genome po

34	97.5	8.1	245	1 VE4_HPV05	P06924 human papil
35	97.5	8.1	1089	1 Y553_HUMAN	Q9UKJ3 homo sapien
36	97	8.0	519	1 IRX4_HUMAN	P78413 homo sapien
37	97	8.0	2333	1 PGCA_CANFA	Q28343 canis fam11
38	96.5	8.0	1324	1 IRS2_HUMAN	Q294H2 mus sapien
39	96.5	8.0	1527	1 CA1H_MOUSE	P39061 mus musculu
40	96.5	8.0	2067	1 NC06_MOUSE	Q9J119 m nuclea r
41	96	7.9	340	1 GBX2_CHICK	Q42230 gallus gall
42	96	7.9	430	1 VSX1_CHICK	Q91A12 gallus gall
43	96	7.9	439	1 XP2_XENLA	P17437 xenopus lae
44	96	7.9	564	1 ARX_MOUSE	O35085 mus musculu
45	96	7.9	1298	1 ICP4_HSV11	P08392 herpes simp

ALIGNMENTS

RESULT 1	ID	Y127_HUMAN	STANDARD;	PRT;	314 AA.
AC	Q14140;				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Hypothetical protein KIAA0127.				
GN	KIAA0127.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Bone marrow;				
RX	MEDLINE=96127530; PubMed=8590280;				
RA	Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;				
RT	"Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line K5-1."				
RL	DNA Res. 2:167-174(1995).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib.ch).				
CC	EMBL: D50917; BAA09476.1; -				
DR	Hypothetical protein.				
KW	SEQUENCE 314 AA; 33896 MW; E43107PC565AAC31 CRC64;				
SQ					
Query Match	17.2%; Score 208.5; DB 1; Length 314;				
Best Local Similarity	27.2%; Pred. No. 3.7e-07;				
Matches	88; Conservative 37; Mismatches 91; Indels 107; Gaps 11;				
QY	1 MLSKGLKKRKEEKEEPLAVDSWMLDPGAAVQAQAPVVA-----SSLPFDLSVTLKHS 56				
DB	1 MLSKGLKKRKEEKEEPLAVDSWMLDPGAAVQAQAPVVA-----SSLPFDLSVTLKHS 56				
QY	57 LQOSEPDLHLVYVNTLRIOASNAAPALPPVSPAPSV-----ADNLTA 105				
DB	52 RLUTPELSLOKTVLINMELRRIQEEKSLRPMFPPSSQPTPEPSDYREAPAFASHIA 111				
QY	106 S-----SDAASASW--ASLLEDLSHIGLSQAPPLADE--GPGG-----142				
DB	112 SPSSHPCDIGSTTPLEACLPASLEDDDDTCTCTSOAMPPTAKISPPALLPEKDSFSS 171				
QY	143 -----RSIGGAPSLGALDLIGP-----160				
DB	172 ALDEIEELCPTSTTEAATATADSVKGTSSGAGTOKLDQPSRADSKLMSLQNFRI 231				
QY	161 --ATGCLLDGLEG-LFEDIDTSMYDNEIMAPAS---EGLKGPEDGPKKEAP-----208				

DB 232 TTSTGFLDRLDLDLFDADIDTSMYDFDCTSSSGTASMAVSAVDLKLTLAPYSSQPV 291
 OY 209 -----ELDEALDYIMDVLCGT 225
 DB 292 TPSOPFKMDLTFELDHIMEVLWS 314

RESULT 2

TEGU_HSV11 STANDARD: PRT: 3164 AA.
 ID TEGU_HSV11
 AC P10220:
 DI 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DF 01-APR-1993 (Rel. 25, Last annotation update)
 DE Large tegument protein (Virion protein UL36).
 GN UL36.
 OS Herpes simplex virus (type 1 / strain 17).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 OX NCBI_TaxID=10299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88274327; PubMed=2839594;
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
 RA McNaughton D., Perry L.J., Scott J.E., Taylor P.;
 RT "The complete DNA sequence of the long unique region in the genome of
 RT herpes simplex virus type 1.";
 RL J. Gen. Virol. 69:1531-1574(1988).
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
 CC -1- FUNCTION: TEGUMENT PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
 CC ENV-1 24, ENV BPFL1, HVS-1 64, VZV 22, AND HCMV UL48.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X14112; CA32311.1; -;
 DR PIR: I30085; WMBEH6.
 DR InterPro: IPR006928; Herpes_teg_N.
 DR InterPro: IPR005210; Herpes_UL36.
 DR Pfam: PF04843; Herpes_teg_N; 1.
 DR Pfam: PF03586; Herpes_UL36; 1.
 DR Repeat.
 KW DOMAIN
 FT SEQUENCE 2911 2980 35 X 2 AA TANDEM REPEATS OF P-Q.
 SQ SEQUENCE 3164 AA; 335857 MW; CC5D31FF4F9FE3F4 CRC64;

Query Match 9.9%; Score 120; DB 1; Length 3164;
 Best Local Similarity 26.7%; Pred. No. 3.2;
 Matches 65; Conservative 23; Mismatches 107; Indels 48; Gaps 10;

OY 21 VDSNMWDPGHAANAQAAPPAVASSSLFDLSYKLHHSQOSRPDLRHLYVNTLRIOAS 80
 DB 2396 IETVAVRLHADLPNENACLAQDLPRLSALIAERPLAKGP-----CLVL-----DIS 2445
 OY 81 MAPAAALPPVSPPAASV-----ADNLASS-----DAALSAMASLLE 120
 DB 2446 MTPVAVLMENDDPPGPPVRRVVGSEATEELPFVAGGEVLAASATDEPFLARAILGPPF 2505
 OY 121 DLSHIEGLSQAPPLADEGPPGSGISGAAPSLGALD-----LGPAN--GCLLDGLEG 172
 DB 2506 DASLISELFGHGVYGRAPDQSPVNPPTPGFVDLVAGSGSLAPTLFTDAVPG 2565
 OY 173 LFEDIDSMYNEMLMAASSEGLAKRPEDGPKKEFAPELDEALDYIMDVLCGT-QALERP 231
 DB 2566 --EVPVPMW--AWIHGLEELIASDGGP-----APLIAADPLSPTAQSVPTSQACAPR 2616
 OY 232 GPG 234
 III

DB 2617 GPG 2619

RESULT 3

SHK1_HUMAN STANDARD: PRT: 2161 AA.
 ID SHK1_HUMAN
 AC Q9Y566; Q9NYW9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DF 28-FEB-2003 (Rel. 41, Last annotation update)
 DE SH3 and multiple ankyrin repeat domain protein 1 (Shank1)
 DE (Somatic protein receptor interacting protein) (SSTR interacting protein)
 DE (SSTRIP).
 GN SHANK1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND INTERACTION WITH SSTR2.
 RC TISSUE=Fetal brain, Hippocampus, and Thalamus;
 RX MEDLINE=20020275; PubMed=1051867;
 RA Zitzer H., Hoencck H.-H., Baechner D., Richter D., Kretzenkamp H.-J.;
 RT "Somatostatin receptor interacting protein defines a novel family of
 RT multidomain proteins present in human and rodent brain.";
 RL J. Biol. Chem. 274:32997-33001(1999).
 RN [2]
 RP REVIEW.
 RX MEDLINE=20267867; PubMed=10806096;
 RA Sheng M., Kim E.;
 RT "The Shank family of scaffold proteins.";
 RL J. Cell Sci. 113:1851-1856(2000).
 CC -1- FUNCTION: Seems to be an adapter protein in the postsynaptic
 CC density (PSD) of excitatory synapses that interconnects receptors
 CC of the postsynaptic membrane including NMDA-type and metabotropic
 CC glutamate receptors via complexes with GRIIP/PSD-95 and Homer,
 CC respectively, and the actin-based cytoskeleton. May play a role in
 CC the structural and functional organization of the dendritic spine
 CC and synaptic junction.
 CC -1- SUBUNIT: May homomultimerize via its SAM domain (By similarity).
 CC Interacts with SSTR2 C-terminus via the PZ domain. Interacts with
 CC SPRY1, Homer-1 and DGAP1/GRIIP isoforms 1 and 2 (By similarity).
 CC Is part of a complex with DLG4/PSD-95 and DGAP1/GRIIP (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
 CC neuronal cells (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=3:
 CC Name=1; Synonyms=A;
 CC IsoId=Q9Y566-1; Sequence=Displayed;
 CC Name=2; Synonyms=B;
 CC IsoId=Q9Y566-2; Sequence=VSP_006069, VSP_006070;
 CC Name=3;
 CC IsoId=Q9Y566-3; Sequence=VSP_006071;
 CC -1- TISSUE SPECIFICITY: Expressed in brain particularly in the
 CC amygdala, hippocampus, substantia nigra and thalamus. Isoform 2
 CC seems to be expressed ubiquitously.
 CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.
 CC -1- SIMILARITY: Contains 6 ANK repeats.
 CC -1- SIMILARITY: Contains 1 PZ/DHR domain.
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF163302; AAD45121.1; -;
 DR EMBL: AF226728; AAF35887.1; -;

DR	HSSP; P06241; 1SHF.
DR	Genew; HGNC:15474; SHANK1.
DR	MIM: 604999; -
DR	InterPro; IPR002110; ANK.
DR	InterPro; IPR001478; PDZ.
DR	InterPro; IPR001660; SAM.
DR	InterPro; IPR001452; SH3.
DR	Pfam; PF00023; ank; 7.
DR	Pfam; PF00595; pdz; 1.
DR	Pfam; PF00536; sam; 1.
DR	Pfam; PF00018; sh3; 1.
DR	ProDom; PD000066; sh3; 1.
DR	SMART; SM00248; ANK; 6.
DR	SMART; SM00228; PDZ; 1.
DR	SMART; SM00454; SAM; 1.
DR	SMART; SM00326; SH3; 1.
DR	PROSITE; PS50088; ANK_REPEAT; 3.
DR	PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR	PROSITE; PS50106; PDZ; 1.
DR	PROSITE; PS50105; SAM_DOMAIN; 1.
DR	PROSITE; PS50002; SH3; 1.
KW	ANK repeat; SH3 domain; Repeat; Alternative splicing.
FT	REPEAT 212 245
FT	REPEAT 246 278
FT	REPEAT 279 312
FT	REPEAT 313 345
FT	REPEAT 346 378
FT	REPEAT 379 395
FT	DOMAIN 554 613
FT	DOMAIN 663 757
FT	DOMAIN 2098 2161
FT	DOMAIN 1002 1007
FT	DOMAIN 1014 1019
FT	DOMAIN 1189 1195
FT	DOMAIN 1709 1717
FT	DOMAIN 1844 1854
FT	DOMAIN 1896 1902
FT	DOMAIN 1970 1979
FT	VARSPLIC 1 613
FT	VARSPLIC 614 654
FT	VARSPLIC 646 654
FT	SEQUENCE 2161 AA; 225019 MW; 5FECC969CBBE98701 CRC64;
Query Match	9.8%; Score 118.5; DB 1; Length 2161;
Best Local Similarity	26.7%; Pred. No. 2.7; Indels 55; Gaps 10;
Matches	66; Conservative 20; Mismatches 106;
QY	10 REEEEREPLAVDSWMILDPGHA-----VAADPAVAVASSLFDLSVLKLNHSIQ 59
Db	1535 RASENGRPLV----LRPAPSVDVDEGEFLFVRLPPLEFSNFKPSPESLTPRGPH 159
QY	60 SEEDLRHLVLYVTLLRIQASMAPRAALPYVPSP--PAPSVADNLASSDAALSAMAS 117
Db	1591 PLRP-----TPPATPLPYPVPAVAAPPTLDS-TASSLTSTSYSEVAT 163
QY	118 LLEDLSHIEGLSOAPOPPLADGGPGGRSI-GGAAPSIGALDLGPATGCLLDGEGEL--- 173
Db	1634 L-----TQGSAAPAGDHPHGCPAPAARAARAPGPC---DPPPG--TSGILEVDNR 168
QY	174 -----FEDIDTSWYNDELMAPIASEGLKPGEDEPGKEAEDELDEALDYLMDEVGTQA 227
Db	1682 SSDHPLETITSSASTLSISLAEGGSAGGGGAGAGVASGPELLDTYVAIYLDGQAFGSSS 174
QY	228 LRPPGP 234
Db	1742 TPGPPY 1748

```

RESULT 4
DREB.CHICK STANDARD; PRT; 652 AA.
ID DREB.CHICK
AC P18302; Q91358; Q91359;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Drebrin (Developmentally regulated brain protein).
GN DBN1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
CX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS E1 AND E2).
RC TISSUE=Brain;
RX MEDLINE=89089137; PubMed=3208110;
RA Kojima N., Kato Y., Shiraio T., Odata K.;
RT "Nucleotide sequences of two embryonic drebrins, developmentally
RL regulated brain proteins, and developmental change in their mRNAs.";
RN Brain Res. 464:207-215(1988).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=93368392; PubMed=8361332;
RA Kojima N., Shiraio T., Odata K.;
RT "Molecular cloning of a developmentally regulated brain protein,
RL chicken drebrin A and its expression by alternative splicing of the
RN drebrin gene.";
RL Brain Res. Mol. Brain Res. 19:101-114(1993).
CC -1- FUNCTION: DREBRINS MIGHT PLAY SOME ROLE IN CELL MIGRATION,
CC EXTENSION OF NEURONAL PROCESSES AND PLASTICITY OF DENDRITES,
CC RESPECTIVELY. BINDS F-ACTIN.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=A;
CC IsoId=P18302-1; Sequence=Displayed;
CC Name=E1;
CC IsoId=P18302-2; Sequence=Vsp_004196;
CC Name=E2;
CC IsoId=P18302-3; Sequence=Vsp_004197;
CC -1- TISSUE SPECIFICITY: BRAIN NEURONS.
CC -1- MISCELLANEOUS: DREBRINS ARE CLASSIFIED INTO TWO FORMS OF THE
CC EMBRYONIC TYPE (E1 AND E2) AND ONE FORM OF THE ADULT TYPE (A). THE
CC TIME COURSE OF THEIR APPEARANCE ARE DIFFERENT FROM EACH OTHER.
CC THEIR STRUCTURES ARE CLOSELY RELATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL; M36961; AAA48750.1; -;
DR EMBL; S65296; AAB28012.1; -;
DR EMBL; S65279; AAB28012.1; JOINED.
DR EMBL; S65280; AAB28012.1; JOINED.
DR EMBL; S65281; AAB28012.1; JOINED.
DR EMBL; S65286; AAB28012.1; JOINED.
DR EMBL; S65289; AAB28012.1; JOINED.
DR EMBL; S65290; AAB28012.1; JOINED.
DR EMBL; S65291; AAB28012.1; JOINED.
DR EMBL; S65292; AAB28012.1; JOINED.
DR EMBL; S65294; AAB28012.1; JOINED.
DR EMBL; S65297; AAB28011.2; -;
DR EMBL; S65230; AAB28010.1; -;
DR InterPro; IPR002108; Actindn_cofin.
DR Pfam; PF00241; cofilin_ADF.1.
DR SMART; SM00102; ADF.1.
DR Actin-binding; Neurone; Alternative splicing.

```

FT	DOMAIN	47	50	POLY-GLY.
FT	DOMAIN	332	339	POLY-SER.
FT	DOMAIN	542	552	POLY-GLU.
FT	DOMAIN	643	646	POLY-GLU.
FT	VASPPLOC	317	362	Missing (in isoform E1).
FT	VASPPLOC	307	405	/FTid-VSP_004196.
FT	VASPPLOC	307	405	Missing (in isoform E2).
FT	VASPPLOC	307	405	/FTid-VSP_004197.
FT	VASPPLOC	307	405	E -> EE (IN REF. 1).
FT	CONFLICT	552	552	
FT	SEQUENCE	652 AA;	71535 MW;	8EF7ZCA549833E65 CRC64;
Query Match				
Best Local Similarity		9.3%;	Score 112.5;	DB 1; Length 652;
Matches		63;	Conservative	18; Mismatches
				76; Indels
				99; Gaps
Oy	5 GLKRRKEEEE-	-----	KEPDVAISWILDPGHAIVAQAAPAVASSLSFDLS	49
Dd	388 GEORTRAAEEETPATPKDSPSTQVAPPAATEQHMPPEGPEDKAAPP	-----		437
Oy	50 VLKLHHSIQSEPD	-----	LRLHLVYNVLRLRIQASMAAALPVPSPPAAS	98
Dd	438 -----DEDDPRPAMTAGADVLGV---	---FLESEPSAPA---BPQVETPG		482
Oy	99 VADNL--LASSDAALSAMASLTEDLSHIEIGSQAPDLADGGPRSGTGAAPSIGALD			156
Dd	483 VAEPFLIELMOGDGAPARA-----	TSTPLRDLDTAGRP-----VPPEGT--		521
Oy	157 LLG-----PATGCCLDLGLEGLEDIDTSNYDNELMA---	PASEGLKPEDGPKRE		205
Dd	522 LLLGEDELPEPATPCDAEQ-----HEEVEEEEEEAATGEPHPGLGYEGYQEGP---			573
Oy	206 EAPELDAEAL 215			
Dd	574 EVPTINGEM 583			
<hr/>				
RESULT 5				
M113_HUMAN				
ID	M113_HUMAN	STANDARD:	PRF:	863 AA.
AC	Q8N3F8; Q8N3N8; Q9BVL9; Q9BY92; Q9UH43; Q9UH44; Q9UH45;			
DT	15-SEP-2003 (Rel. 42, Created)			
DT	15-SEP-2003 (Rel. 42, Last sequence update)			
DE	15-SEP-2003 (Rel. 42, Last annotation update)			
CN	MIRAB13 OR KIAI1668.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX	NCB1_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-Uterine adenocarcinoma;			
RA	Pandaitan R.; Zahraoui A.;			
RT	"MIRAB13, a novel effector of Rab13."			
RL	Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.			
RX	[2]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=20057165; PubMed=10591208;			
RA	Dunham I., Hunt L.R., Collins J.E., Bruskiewich R., Beare D.M.,			
RA	Clamp M., Smink L.J., Alnsconough R., Almeida J.P., Babbage A.K.,			
RA	Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,			
RA	Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,			
RA	Burhill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,			
RA	Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,			
RA	Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,			
RA	Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,			
RA	Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,			
RA	Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,			
RA	Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,			
RA	Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,			
RA	Laird G.K., Langford C.F., Leverkusha M.A., Lloyd C., Lloyd D.T.,			
RA	Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.M.,			
RA	Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,			

CC CC
RA Odell C.N., Pavlit R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L.,
RA Scott C.E., Sehna H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swan R.M.,
RA Vadin M., Wall M., Wallis J.M., Whiteley M.N., Wiley D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilmshurst L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Miotsuka S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu Y., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Pang F., Pu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Lon P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chiesse S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozerisky P., Rohlfing T.,
RA Schiet P., Walker C., Wamsley A., Wohlmann P., Popin K., Nelson J.,
RA Koif I., Bedell J.A.K., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel K.S., Shaikh T., Kurahashi H., Salta S., Budarf M.L.,
RA McEwen H.E., Johnson A., Wong A.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M., Dong D.C., Morrow B.E., Edelmann L.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P., Kedra D.,
RA Wilkinson P., Bodenreich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [3]
RP SEQUENCE OF 25-863 FROM N.A.
RC TISSUE-Melanoma:
RA Ansore W., Wirkner U., Mewes H.-W., Well B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 73-863 FROM N.A.
RC TISSUE-Brain:
RA MEDLINE=21156230; PubMed=11258795;
RX Hirotsawa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;
RT "Identification of novel transcribed sequences on human chromosome 22
by expressed sequence tag mapping.";
RL DNA Res. 8:1-9(2001).
RN [5]
RP SEQUENCE OF 514-863 FROM N.A.
RC TISSUE-Chorioncarcinoma:
RA MEDLINE=22388257; PubMed=12477932;
RX Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Butlow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshilyukki S., Carninci P., Prange C.,
RA Rana S.S., Lequelello N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., Moegen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wolzy K.C., Hale S., Garcia A.M., Gay L.J., Hulky S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska T., Smallus D.E.,
RA Schenker A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SUBMIT: Blinds to Rabits.
CC - SIMILARITY: Contains 1 calpain-in-homology (CH) domain.
CC - SIMILARITY: Contains 1 LIM zinc-binding domain.
CC - CAUTION: Ref.3 (CAD39036) sequence differs from that shown due to
frame shifts in positions 486 and 507.


```

CC CC      Isoid-Q96NM4-1; Sequence-Displayed;
CC CC      Name=2;
CC CC      Isoid-Q96NM4-2; Sequence-VSP_002187;
CC CC      Note-No experimental confirmation available;
CC CC      -1- SIMILARITY: Contains 1 HMG box domain.
CC CC      -1- CAUTION: It is uncertain whether Met-1 or Met-52 is the initiator.
CC CC      -1- CAUTION: Ref.1 (CAC36288) sequence differs from that shown due to
CC CC      erroneous gene model prediction.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
DR DR      EMBL; AK051535; BAB70860.1; -
DR DR      EMBL; AL034419; CAD43476.1; -
DR DR      EMBL; AL034419; CAD43477.1; -
DR DR      EMBL; AL121887; CAC36288.1; ALT_SEQ.
DR DR      EMBL; BC007636; AAH07636.1; -
DR DR      Genew; HGNC:16095; C20orf100.
DR DR      InterPro; IPR000910; HMG_12_Box.
DR DR      Pfam; PF00505; HMG-box; 1.
DR DR      PROSITE; PS50118; HMG_BOX_2; 1.
DR DR      Nuclear protein; DNA-binding; Alternative splicing.
FT FT      DNA_BIND 255 323
FT FT      DOMAIN 372 456
FT FT      VASNPIC 302 302
FT FT      PRO-RICH.
FT FT      ISOFORM 2.
FT FT      /FTID-VSP_002187.
FT FT      D -> N (IN REF. 1).
SQ SEQUENCE 488 AA; 51604 MW; 687FD144CF30731A CRC64;

Query Match
Best Local Similarity 26.0%; Score 108.5; DB 1; Length 488;
Matches 58; Conservative 28; Mismatches 90; Indels 47; Gaps 10;

QY 7 KRRKEEERKEPLAVDSMTLDPRGAAYAAQAPPAVASSSLDLYL-----KLHHSLOOS 60
DB 247 KKKKDPNEQKPVSAVALFFRDQALIKGNP---SATFGVSKIVASMSDLSDEQKOS 303
QY 61 EPLRLHVLVYVNTLRIRIQASMAPPAALPPVSPPAASVA-----DNILASSDAALSASM 115
DB 304 SPD-----QGETKSTQAN-PPAKMLPPQOPMYAMPGLASFLTPSDLOAFRSGASPASL 355
QY 116 ASLEDLSHIEGLSQAPPLADEGPPRSIGGAPSL-GALDLDGPATGCLLDGLEGLE 174
DB 356 ARTLGSSSLPGLSASPPP-----PPSPFL---SPTLHQDLSLPPHAGALLS----- 400
QY 175 EDIDTSMYDNEMLAPASEGLK-----PGPEDGFGKEEAP 208
DB 401 PAVSMSPAPQPPVLPPTPMALQVOLAMSPSPQDPFHISEFP 443

RESULT 7
FXGA_CHICK STANDARD; PRT; 440 AA.
AC 098937;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Forkhead box protein G1A (Forkhead-related protein FKHL2)
DE (Transcription factor BF-2) (Brain factor 2) (BF2) (CBF-2) (T-14-6).
GN FOXG1A OR FKHL2 OR HFBF2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_Taxid=9031;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn; TISSUE=Retina;
RX MEDLINE=96338226; PubMed=8757134;
RA Yuasa J., Hirano S., Yamagata M., Noda M.;
RT "Visual projection map specified by topographic expression of
RT transcription factors in the retina.";
RL Nature 382:632-635(1996).
CC CC      -1- FUNCTION: MAY DETERMINE THE NASOTEMPORAL AXIS OF THE RETINA, AND
CC CC      CONSEQUENTLY SPECIFY THE TOPOGRAPHICAL PROJECTION OF THE RETINAL
CC CC      GANGLION-CELL AXONS TO THE TECTUM BY CONTROLLING EXPRESSION OF
CC CC      THEIR TARGET GENES.
CC CC      -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC CC      -1- TISSUE SPECIFICITY: Retina.
CC CC      -1- DEVELOPMENTAL STAGE: CAN BE DETECTED IN REGIONS INCLUDING
CC CC      PRIMITIVE RETINA AND NEUROEPITHELIUM BY EMBRYONIC DAY 2 (E2). AT
CC CC      E3, EXPRESSED IN THE TEMPORAL RETINA AND ASSOCIATED PIGMENT
CC CC      EPITHELIUM AS WELL AS IN PART OF THE DIENCEPHALON, AND AT E7 IS
CC CC      EXPRESSED IN RETINAL GANGLION CELLS. LEVELS BEGIN TO DECLINE FROM
CC CC      E4 AND ALMOST DISAPPEAR BY E10.
CC CC      -1- SIMILARITY: Contains 1 fork-head domain.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
DR DR      EMBL; U47276; AAB08467.1; -
DR DR      PIR; S71795; S71795.
DR DR      HSSP; Q63245; ZHPH.
DR DR      InterPro; IPR001766; TF_Fork_head.
DR DR      Pfam; PF00250; Fork_head; 1.
DR DR      PRINTS; PR00053; FORKHEAD.
DR DR      PRODOM; PD000425; TF_Fork_head; 1.
DR DR      SMART; SM00339; FH.1.
DR DR      PROSITE; PS00657; FORK_HEAD_1; 1.
DR DR      PROSITE; PS00658; FORK_HEAD_2; 1.
DR DR      PROSITE; PS50039; FORK_HEAD_3; 1.
DR DR      Transcription regulation; DNA-binding; Nuclear protein;
DR DR      developmental protein.
FT FT      DOMAIN 39 43
FT FT      DOMAIN 49 58
FT FT      DOMAIN 77 88
FT FT      DOMAIN 109 112
FT FT      DOMAIN 113 137
FT FT      DNA_BIND 143 237
FT FT      DOMAIN 321 330
FT FT      POLY-PRO.
SQ SEQUENCE 440 AA; 44669 MW; 8226C2E1E103A48 CRC64;

Query Match
Best Local Similarity 32.3%; Score 107.5; DB 1; Length 440;
Matches 40; Conservative 12; Mismatches 45; Indels 27; Gaps 6;

QY 82 APAALPPVSPPAASVADNLLASSDA-----LASMASLEDLSHIEGLSQAP 132
DB 315 APAALPPPPPPPPRRAPLPPAELARTPPGYPPLGLPALAASLHAAPSGAAYAR 374
QY 133 OPTADEGPPRSIGGAPSLGALDLDGPATG---CLLDG-LEGLEDITSMYDNEML 187
DB 375 SPFSIESITG---GGPFGIGA-----GPAFGAGSCASGAGATGLSRISGSL----- 421
QY 188 APAS 191
DB 422 APAA 425

RESULT 8
VP40_HCMVA STANDARD; PRT; 708 AA.
AC 16753; 069030;
DT 01-AUG-1990 (Rel. 15, Created)

```


DB 365 --LANPHGVLYPKAFSLGASHANVPMTPGAVAPPSAPDLPLSPASTYGAIV 422
 QY 210 LDEAL-----DYL 218
 DB 423 VGYPDLAARHAFADY 437

RESULT 9
 DUS8_MOUSE STANDARD; PRT; 663 AA.
 AC 009112;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dual specificity protein phosphatase 8 (EC 3.1.3.48) (EC 3.1.3.16)
 DE (Neuronal tyrosine threonine phosphatase 1).
 GN DUSP8 OR NTPPL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96311565; PubMed=8733137;
 RA Theodosiou A.N., Rodrigues N.R., Nesbitt M.A., Ambrose H.J.,
 RA Paterson H., McLeilan-Arnold E., Boyd Y., Leversha M.A., Owen N.,
 RA Blake D.J., Ashworth A., Davies K.E.;
 RT "A member of the MAP kinase phosphatase gene family in mouse
 RT containing a complex trinucleotide repeat in the coding region.";
 RT Hum. Mol. Genet. 5:675-684(1996).
 CC -1- FUNCTION: THIS PROTEIN SHOWS BOTH ACTIVITY TOWARD TYROSINE-PROTEIN
 CC PHOSPHATE AS WELL AS WITH SERINE/THREONINE-PROTEIN PHOSPHATE (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
 CC phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN AND LUNG.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY. DUAL SPECIFICITY SUBFAMILY.
 CC -1- SIMILARITY: Contains 1 rhodanese domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X95518; CAA64772.1; -;
 DR HSSP; Q16828; IMRP.
 DR MGD; MGI:106826; Dusp8.
 DR InterPro; IPR000340; DS_phosphatase.
 DR InterPro; IPR001763; Rhodanese-like.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR Pfam; PF00782; DSpC; 1.
 DR SMART; SM00195; DSpC; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PS50206; RHODANES_3; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
 KW Hydroxylase; Nuclear protein.
 FT DOMAIN 23 138 RHODANES.
 FT 162 432 PROTEIN-TYROSINE PHOSPHATASE.
 FT 452 459 POLY-ARG.
 FT 555 558 POLY-SER.
 FT 559 576 POLY-GLY.
 FT 577 600 POLY-SER.
 FT DOMAIN 311 552 PRO-RICH.

FT ACT SITE 246 246 BY SIMILARITY.
 SQ SEQUENCE 663 AA: 68847 MW; 416F429A12C1FA7C CRC64;
 Query Match 8.8%; Score 106.5; DB 1; Length 663;
 Best Match Similarity 26.6%; Pred. No. 4.5;
 Matches 53; Conservative 21; Mismatches 66; Indels 59; Gaps 10;

QY 27 DPGHAANAQAPPAVASSSLFDLSYLKLHHSLSQOSEPDRLHVLVYNT-----LRRIOASMA 82
 DB 270 DDATRFVVDORRPSISPNFNLGQLLEVERSLK-----LLAALQTGPHLGTPEPLMG 321
 QY 83 PAALPVPSPPAASVADNLIASDALSSMASLSDLSHIEGLSOAPQPLADECPG 142
 DB 322 PAAG---IPLPRLPSTSESAATGSEAAATAAR-----EGSP- 354
 QY 143 RSIGGAAPSLGALDLGPAFGCLDDGLEGFEDIDTSMYDNEI-----WAPAS 191
 DB 355 -SAGGDAP-----IPSTAPATSA-LQOGLRGHLHSSDRLODTNRKRFSLDIKSAVAPSR 408
 QY 192 EGLKPPEDGPKKEAPEL 210
 DB 409 RPDFPGPD-PG--EAPKL 424

RESULT 10
 CNG4_HUMAN STANDARD; PRT; 909 AA.
 AC 014028; Q14029;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cyclic-nucleotide-gated cation channel 4 (CNG channel 4) (CNG-4)
 DE (CNG4) (Cyclic nucleotide-gated cation channel modulatory subunit).
 GN CNG4 OR CNGC4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS RCNC2A AND RCNC2B).
 RC TISSUE=Retina;
 RX MEDLINE=93226050; PubMed=7682292;
 RA Chen T.Y., Peng Y.-W., Dhallan R.S., Ahmed B., Reed R.R., Yau K.-W.;
 RT "A new subunit of the cyclic nucleotide-gated cation channel in
 RT retinal rods.";
 RL Nature 362:764-767(1993).
 CC -1- SUBUNIT: HETEROOLIGOMERIC COMPLEX WITH CNG1.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: Named isoforms-2;
 CC Event-Alternative splicing; Named isoforms-2;
 CC Name=RCNC2B;
 CC IsoId=Q14028-1; Sequence=Displayed;
 CC Name=RCNC2A;
 CC IsoId=Q14028-2; Sequence=VSP_001110;
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L15296; AAA65620.1; -;
 DR EMBL; L15297; AAA65619.1; -;
 DR MIM; 600724; -;
 DR GO; GO:0017071; C:intracellular cyclic nucleotide activated c. . .; NAS.
 DR GO; GO:0005222; F:intracellular cAMP activated cation channel. . .; NAS.
 DR InterPro; IPR000595; cAMP_binding.
 DR InterPro; IPR005821; Ion_trans.
 DR InterPro; IPR001622; K-channel_pore.

DR pfam: PF00027; CNMP_binding; 1.
 DR SMART: SM00100; CNMP; 1.
 DR PROSITE: PS00888; CNMP_BINDING_1; 1.
 DR PROSITE: PS00889; CNMP_BINDING_2; 1.
 DR PROSITE: PS50042; CNMP_BINDING_3; 1.
 KW Ionic channel; Ion transport; CAMP-binding; Transmembrane;
 KW Multigene family; Alternative splicing.
 FT DOMAIN 1 314
 FT TRANSMEM 315 333
 FT DOMAIN 334 347
 FT TRANSMEM 348 366
 FT DOMAIN 367 391
 FT TRANSMEM 392 411
 FT DOMAIN 412 448
 FT TRANSMEM 449 471
 FT DOMAIN 472 515
 FT TRANSMEM 516 535
 FT DOMAIN 536 619
 FT TRANSMEM 620 640
 FT DOMAIN 641 909
 FT NP_BIND 628 767
 FT BINDING 688 688
 FT BINDING 700 700
 FT VARSPLIC 1 286
 FT SEQUENCE 909 AA; 102285 MW; DCOE754336BEDD CRC64;
 Query Match 8.8%; Score 106.5; DB 1; Length 909;
 Best Local Similarity 21.7%; Pred. No. 6.4;
 Matches 51; Conservative 34; Mismatches 89; Indels 61; Gaps 10;
 QY 11 EEEEREPLAVDSWMLPGHAAVAOAPPAVASSLEFLSVLKHSILQOSEPLRLHLYV 70
 DB 88 EEEAEKEP---QDM-----AEKKEPEAEAEKASGVAATKQHPVEVEDD----- 131
 QY 71 VNTLRITQASMAPAALPPVPS---PPAAPVADNILLASDAALS--ASMASILEDLSHI 125
 DB 132 -----ADSCPLAMEENPSTVLPSPSPAKSDTLIVSSASGTHKKLPSSEDDAEEL 183
 QY 126 EGLSOAPOPPLADGCP-----GRSIGGAAPSLGALDLGPRATGCLLDGLEGF---- 174
 DB 184 KALSPAESPVVANSDFTPPKDQDQRAASTA-----TNSAINDRIQELVKLEK 234
 QY 175 ---EDIDTSMYDNELW-----APASEGLKPGEDGKGEAPLDEAELDYL 218
 DB 235 ERTEKYEKELIDPDVTSDEESPKRPAKAPAPAPDTKFAAEF-PVEEHYCDML 288
 RESULT 11
 ODP2_RHIME STANDARD; PRT: 447 AA.
 ID ODP2_RHIME
 AC O9R9N3;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase
 DE complex (EC 2.3.1.12) (E2).
 GN PDHC OR PDH OR R01447 OR SMC01032.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RCR2011 / SU47;
 RX MEDLINE=20255031; PubMed=10796014;
 RA Cabanes D., Boistard P., Batut J.;
 RT "Symbiotic induction of pyruvate dehydrogenase genes from
 RT Sinorhizobium meliloti.";
 RL Mol. Plant Microbe Interact. 13:483-493(2000).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaire V., Masuy D.,
 RA Pohl T., Portetelle D., Puehler A., Punelle B., Ramsperger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallbert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 CC -1- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
 CC CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
 CC COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
 CC DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
 CC (E3) (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide = CoA + S-
 CC acetylhydrolipoamide.
 CC -1- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL
 CC COFACTOR (BY SIMILARITY).
 CC -1- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL
 CC SYMMETRY (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.
 CC -1- SIMILARITY: Contains 1 lipoyl-binding domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AF190792; AAF04589.1; -
 CC EMBL: AL591787; CAC46026.1; -
 DR HSSP: P07016; 1BLU.
 DR InterPro: IPR001078; 2Oxoacid_dh.
 DR InterPro: IPR006257; Acet_long.
 DR InterPro: IPR000089; Biotin_lipoyl.
 DR InterPro: IPR004167; E3 binding.
 DR InterPro: IPR003016; Lipoyl.
 DR Pfam: PF00198; 2-oxoacid_dh; 1.
 DR Pfam: PF00364; biotin_lipoyl; 1.
 DR Pfam: PF02817; e3_binding; 1.
 DR ProDom: PD001115; 2Oxoacid_dh; 1.
 DR TIGRFAMS: TIGR01349; PDHac_trf_mito; 1.
 DR PROSITE: PS00189; LIPOYL; 1.
 DR GlycoLysis: Transferase; Acyltransferase; Lipoyl; Complete proteome.
 KW BINDING 43 43 LIPOYL (BY SIMILARITY).
 FT ACT_SITE 420 420 POTENTIAL.
 FT SEQUENCE 447 AA; 46140 MW; 48B1CAC4E2FDC2AC CRC64;
 Query Match 8.7%; Score 105; DB 1; Length 447;
 Best Local Similarity 31.1%; Pred. No. 3.7;
 Matches 52; Conservative 14; Mismatches 51; Indels 50; Gaps 9;
 QY 79 ASMAPAALPPVSPPAASVAD-----NLLASDAALSMSASILEDLSHIGLSGAPPP 134
 DB 113 AAPAPAAAPQAAAPASAPADGCKRTFFSSPIARRLAKKEGI--DLAING----- 163
 QY 135 LADEGPGGSI-----GGAAPSLGALDLGAPATGCLLDGLEGLEDIDTSMYDNE 185
 DB 164 ---SGPHGVVKKVDVETAVSGAARPAKPA-PAAPAPATL-----AKGNSDEAVLKLEF-- 213
 QY 186 LMAPASEGLKPGEDGKGEAPLDEA-----ELDYLM 219
 DB 214 ---PGSYELV--PHDGMRTIAKRLVESKOTIPHFYVSVDCBLDALM 255
 RESULT 12
 FNM2_MOUSE
 ID FNM2_MOUSE
 AC O9JL04;
 STANDARD; PRT: 1567 AA.

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Formin 2.
 GN FMN2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20245324; PubMed=10781961;
 RA Leader B., Leder P.;
 RT "Formin-2, a novel formin homology protein of the cappuccino
 RT subfamily, is highly expressed in the developing and adult central
 RT nervous system."
 RL Mech. Dev. 93:221-231(2000).
 CC -1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN THE DEVELOPING
 CC AND MATURE CENTRAL NERVOUS SYSTEM.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS AT EMBRYONIC DAY 9. 5 IN
 CC THE DEVELOPING SPINAL CORD AND BRAIN STRUCTURES AND CONTINUES IN
 CC NEONATAL AND ADULT BRAIN STRUCTURES INCLUDING THE OLFACTORY BULB,
 CC COREX, THALAMUS, HYPOTHALAMUS, HIPPOCAMPUS AND CEREBELLUM.
 CC -1- SIMILARITY: Contains 1 formin homology 1 (FH1) domain.
 CC -1- SIMILARITY: Contains 1 formin homology 2 (FH2) domain.
 CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. CAPPUCCINO
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF218940; AAF72883.1; -
 DR MGD: MGI:1859252; Fmn2.
 DR InterPro: IPR003104; FH2.
 DR InterPro: IPR001265; Formin.
 DR Pfam: PF02181; FH2; 1.
 DR PRINTS: PRO00828; FORMIN.
 DR SMART: SM00498; FH2; 1.
 KM Developmental protein; Repeat: Coiled coil.
 FT DOMAIN 643 683
 FT 735 1113
 FT DOMAIN 919 1039
 FT 919 929
 FT REPEAT 930 940
 FT REPEAT 941 951
 FT REPEAT 952 962
 FT REPEAT 963 973
 FT REPEAT 974 984
 FT REPEAT 985 995
 FT REPEAT 996 1006
 FT REPEAT 1007 1017
 FT REPEAT 1018 1028
 FT REPEAT 1029 1039
 FT REPEAT 1128 1132
 FT DOMAIN 1408 1444
 FT 1408 1444
 FT DOMAIN 48 55
 FT 202 207
 FT DOMAIN 797 801
 FT 861 864
 FT DOMAIN 908 917
 FT 922 928
 FT DOMAIN 933 939
 FT 944 950
 FT DOMAIN 955 961
 FT 966 972
 FT DOMAIN 977 983
 FT 983 983

FT DOMAIN 988 994
 FT 999 1005
 FT DOMAIN 1010 1016
 FT 1021 1027
 FT DOMAIN 1032 1038
 FT 1043 1046
 FT DOMAIN 1054 1060
 FT 1065 1072
 FT DOMAIN 1077 1080
 FT 1080 1080
 SQ SEQUENCE 1567 AA; 166268 MW; 8F273B1C88505944 CRC64;
 Query Match 8.6%; Score 103.5; DB 1; Length 1567;
 Best Local Similarity 20.4%; Pred. No. 18;
 Matches 64; Conservative 33; Mismatches 100; Indels 117; Gaps 12;
 QY 4 GGLKRRREEEKEPLAVDSWMLDPGHAANQAP----- 37
 DB 84 KGACDSRDVDSQALPTGE--LDSAHSHVTKPTDLSABETGLSDTECADPREVTHPG 141
 QY 38 ---PAVA-----SSLEPDLVYKLIHSHIQOSE--PDLRHLYLV 70
 DB 142 ASRPAAAGVGIQATAEDELTAAGADGCGRTSSGSDTDYFSHATGEQEDLSDIQAIRL 201
 QY 71 VNTLRR---IQASMAAALPPV--PSPAPASVDNLASSDA-----L 111
 DB 202 QQQQQQKLLLDSEPPAA--PPTAISPPGAFGLGDPLDPRSEAEKDTVQALPVPDL 259
 QY 112 SASNASLED-----LSHIEGLSQAPPLADE--GPPGRSIGAPASIGALDLGPA 161
 DB 260 PETRSLVPEHPSPSGSHLTSETGYATAPASAVDLSLSPAFPTPEAGPGGAAGVPVAG 319
 QY 162 TGCLLDGLBGLFEDIDTSMTDNEIMAPASGLKRGPDGCKEAPELDAEILDYLMDV 221
 DB 320 TGDDEDECEDEAFED-----APRGSPEGEWPEMEBA----- 351
 QY 222 LVGTQALERPFGPG 235
 DB 352 ---SORLKEPPEG 362
 RESULT 13
 A180_MOUSE STANDARD; PRT; 901 AA.
 AC 061548; 061547;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Clathrin coat assembly protein APl80 (Clathrin coat associated protein
 DE APl80) (91 kDa synaptosomal-associated protein) (phosphoprotein Fl-
 DE 20).
 GN SNAP91.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RX MEDLINE=92300439; PubMed=1607933;
 RA Zhou S., Sousa R., Tannery N.H., Lafer E.M.;
 RT "Characterization of a novel synapse-specific protein. II. cDNA
 RT cloning and sequence analysis of the Fl-20 protein."
 RL J. Neurosci. 12:2144-2155(1992).
 CC -1- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES WHICH
 CC LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-ASSOCIATED
 CC PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE CYTOPLASMIC
 CC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION AND
 CC CONCENTRATION. BINDING OF APl80 TO CLATHRIN TRISKELIA INDUCES
 CC THEIR ASSEMBLY INTO 60-70 NM COATS.
 CC -1- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE
 CC CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=long;

CC IsoId-Q61548-1; Sequence-Displayed;
CC Name-Short;
CC IsoId-Q61548-2; Sequence-VSP_000172;
CC TISSUE SPECIFICITY: BRAIN. ASSOCIATED WITH THE SYNAPSES.
CC -1- DEVELOPMENTAL STAGE: DEVELOPMENTALLY REGULATED IN A PATTERN
CC COINCIDENT WITH ACTIVE SYNAPTONESTIS AND SYNAPTIC MATURATION.
CC -1- DOMAIN: POSSESSES A THREE DOMAIN STRUCTURE: THE N-TERMINAL 300
CC RESIDUES HARBOUR A CLATHRIN BINDING SITE, AN ACIDIC MIDDLE DOMAIN
CC 450 RESIDUES, INTERRUPTED BY AN ALA-RICH SEGMENT, AND THE C-
CC TERMINAL DOMAIN (166 RESIDUES).
CC -1- PTM: Phosphorylated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M83985; AAA37587.1; -
CC EMBL; M83985; AAA37586.1; -
CC PIR; A44825; A44825.
CC MGD; MGI109132; Snap91.
CC InterPro; IPR001026; ENTH.1.
CC SMART; SM00273; ENTH.1.
CC Coated pits; Alternative splicing; Phosphorylation.
CC DOMAIN 410 413 POLY-THR.
CC FT DOMAIN 535 539 POLY-ALA.
CC FT DOMAIN 547 550 POLY-ALA.
CC FT DOMAIN 659 664 POLY-SER.
CC FT DOMAIN 704 710 POLY-SER.
CC FT VARSPLIC 715 719 Missing (in isoform Short).
CC /FTId-VSP_000172.
SO SEQUENCE 901 AA; 91851 MW; 24498FBACEBDBB1 CRC64;

Query Match 8.5%; Score 102.5; DB 1; Length 901;
Best Local Similarity 21.2%; Pred. No. 12;
Matches 53; Conservative 25; Mismatches 93; Indels 79; Gaps 9;

QY 32 AVAQAAPPAAVSSSLFSLVSLKLNHSLQOQSEPDRLHLYLVNLTIRKIQASMAPAAP-- 89
DB 518 AFSPAATVAATAATTTA-----AAAATTATTSAAATTAAPAL 559
QY 90 -----VSPPAAPSV-----ADNLASSDAALSASMSLLEDLSHIGLSQ 130
DB 560 DIRGDLFDSAPFAAPKPDAAPSIDLFGTDAFSSPPRGASPPRESSLTRADLSVAF 618
QY 131 APOPLADEGPPGRSGAAPSIGALDILGPGATGCLLDGLGLELFD-----IDTSM--- 181
DB 619 AFSPAATASP-----AKAESGVIDLFGDAFG-----SGAETGPAPQAVASSSASA 665
QY 182 -----YDNLMAFASGLKRPED-----GPKKEAPELDEALDYLMVLYG 224
DB 666 DLIAAGSGFMASTTPTPAONNLQPSFEAFAGTTPSTSSSSSPDPVSFGLDILMP 725
QY 225 TQALRPPGP 234
DB 726 TMAPSGOPAP 735

RESULT 14
SRCA_RABIT
ID SRCA_RABIT STANDARD; PRT; 908 AA.
AC P13666;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sarcalumenin precursor.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=89345602; PubMed=2762314;
RA Leberer E., Charuk J.H.M., Green N.M., MacLennan D.H.;
RT "Molecular cloning and expression of cDNA encoding a luminal calcium
RT binding glycoprotein from sarcoplasmic reticulum."
RL Proc. Natl. Acad. Sci. U.S.A. 86:6047-6051(1989).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=89123480; PubMed=2521635;
RA Leberer E., Charuk J.H.M., Clarke D.M., Green N.M., Zubrycka-Gaarn E.,
RA MacLennan D.H.;
RT "Molecular cloning and expression of cDNA encoding the 53,000-dalton
RT glycoprotein of rabbit skeletal muscle sarcoplasmic reticulum."
RL J. Biol. Chem. 264:3484-3493(1989).
CC -1- FUNCTION: PERAPAS INVOLVED IN THE REGULATION OF CALCIUM TRANSPORT.
CC -1- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN. ASSOCIATED
CC THROUGH CA(2+) WITH THE MEMBRANE.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=160 kDa;
CC IsoId-P13666-1; Sequence-Displayed;
CC Name=2; Synonyms=58 kDa;
CC IsoId-P13666-2; Sequence-VSP_004429;
CC Note-No experimental confirmation available;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M25750; AAA31189.1; -
CC DR EMBL; J04480; AAA60730.1; -
CC PIR; A33280; A33280.
CC KX Calcium-binding; Glycoprotein; Signal; Alternative splicing.
CC FT SIGMNL 1 19
CC FT CHAIN 20 908
CC FT DOMAIN 20 457 SARCALUMENIN.
CC FT CARBOHYD 102 102 ACIDIC DOMAIN, PROBABLY BINDS CALCIUM.
CC FT CARBOHYD 716 716 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT VARSPLIC 20 455 Missing (in isoform 2).
CC /FTId-VSP_004429.
CC CONFLICT 474 474 Q -> E (in ref. 2).
SO SEQUENCE 908 AA; 97920 MW; A48CAA221AE1A18B CRC64;

Query Match 8.5%; Score 102.5; DB 1; Length 908;
Best Local Similarity 26.2%; Pred. No. 12;
Matches 63; Conservative 24; Mismatches 104; Indels 49; Gaps 11;

QY 32 AVAQAAPPAAVSSSLFSLVSLKLNHSLQOQSEPDRLHLYLVNLTIRKIQASMAPAAP-- 89
DB 19 AELQVP-----ASGTEDEVNLTENHFSAGDASLEEKRALYDAAPDENLHLYPGREA 75
QY 78 QASMAPAAPALP-VPSPPAAPSVADNLASSDAALSA-----SNASLLEDL 122
DB 76 ESAERTTAGAPPAAPGSEPEAS--LPNASFTESAPPGDAGTPREBOGPAAPASALPPG 132
QY 123 SHIEGLSOAPOPAD-EGPPGRSIGGAPSIGAL--DLGPGATGCLLDGLGLELFDIDT 179
DB 133 GDEGSLQERQELSSGEGFEAAGLITSEGAASGEMOGAGGVPPAEAGVGLDSPV 192
QY 180 SWYDNLMAFASGLKRPED-----GPK-EEAPELDEALDYLMVLYGQALERP 231
DB 193 OGAAETAPPEASGAPSSSEDEQIHTLEEGKSGSPDPGPFELDQTPD---GASAGEEP 249

RESULT 15
CALH_HUMAN

AC ID CAI9060; G9URK38; STANDARD; PRT; 1516 AA.
AD DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].
GN COL18A1.
OS Homo sapiens (Human).
OC Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98164096; PubMed=9503365;
RA Saarela J., Ylikarppa R., Rein M., Purmonen S., Pihlajaniemi T.;
RT "Complete primary structure of two variant forms of human XVII
RT collagen and tissue-specific differences in the expression of the
RL corresponding transcripts";
RL Matrix Biol. 16:319-328(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Itohi Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Tauden S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumfi K., Lehmann R., Patterson D.,
RA Rosenthal A., Rump A., Schillhaber M., Schudy A., Zimmermann W.,
RA Reichenhal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Saeki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Mimosima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riessmann L., Dagand E.,
RA Weinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehnach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE OF 834-1516 FROM N.A.
RX MEDLINE=94245237; PubMed=8186291;
RA Oh S.P., Waxman M.L., Seidlin M.F., Cheng S., Knoll J.H., Timmons S.,
RA Olsen B.R.;
RT "Cloning of cDNA and genomic DNA encoding human type XVII collagen
RT and localization of the alpha 1(XVII) collagen gene to mouse
RL chromosome 10 and human chromosome 21.";
RL Genomics 19:494-499(1994).
RN [4]
RP SEQUENCE OF 1334-1516 FROM N.A.
RX MEDLINE=20400145; PubMed=10942434;
RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RT "Cloning and expression of human endostatin gene in Escherichia
RT coli.";
RL Submitted (SEP-1999) to the EMBL/genbank/DBD databases.
RN [5]
RP INVOLVEMENT IN KNOBLOCH SYNDROME.
RX MEDLINE=20400145; PubMed=10942434;
RA Settle A.L., Sossi V., Camargo A.A., Zatz M., Brabe C.,
RA Passos-Bueno M.R.;
RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RT and tumor growth, plays a critical role in the maintenance of retinal
RT structure and in neural tube closure.";
RL Hum. Mol. Genet. 9:2051-2058(2000).
RN [6]
RP VARIANT ASN-1437.
RX MEDLINE=21518361; PubMed=11606364;
RA Tognetti P., Suzuki O., Godot P.H., Alves V.A., Settle A.L.,
RA Zorick T., Soares F., Camargo A.A., Moreira E.S., di Loreto C.,
RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT for the development of prostatic adenocarcinoma.";
RL Cancer Res. 61:7375-7378(2001).
RN [7]
RP -1- FUNCTION: COL18A1 PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
CC -1- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL

CC		PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH FACTOR SIGNALING.
CC	-1	ALTERNATIVE PRODUCTS:
CC		Event-Alternative splicing; Named Isoforms=2;
CC	Name=Long;	Synonyms=NC_493;
CC	Isoid=p39060-1;	Sequence=DsDisplayed;
CC	Name=Short;	Synonyms=NC1-303;
CC	Isoid=p39060-2;	Sequence=VSP_001155, VSP_001156.
CC	-1	TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS IN LIVER, LONG AND KIDNEY.
CC	-1	PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
CC	-1	POLYMORPHISM: There is an association between a polymorphism in position 1437 and prostate cancer. Heterozygous Asn-1437 individuals have a 2.5 times increased chance of developing prostate cancer as compared with homozygous Asp-1437 individuals.
CC	-1	DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (NNO) [MM:567750]. An autosomal recessive disorder defined by the occurrence of high myopia, vitreoretinal degeneration with retinal detachment, macular abnormalities and occipital encephalocele.
CC	-1	SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC		-----
DR	EMBL;	AF018081; AAC39658.1; -
DR	EMBL;	AF018082; AAC39659.1; -
DR	EMBL;	AL163302; CAB90482.1; -
DR	EMBL;	L22548; AAA51864.1; -
DR	EMBL;	AF184060; AAF01310.1; ALT_INIT.
DR	PDB;	1BN1; O2-DEC-98.
DR	GlycoSuiteDB;	P39060; -
DR	GeneW;	HGNM:2195; COL18A1.
DR	MIM;	120328; -
DR	MIM;	267750; -
DR	GO;	GO:0005581; C:collagen; TAS.
DR	GO;	GO:0008181; F:tumor suppressor; TAS.
DR	GO;	GO:0007397; P:histogenesis and organogenesis; TAS.
DR	GO;	GO:0008285; P:negative regulation of cell proliferation; TAS.
DR	GO;	GO:0007601; P:vision; TAS.
DR	InterPro;	IPIR000087; Collagen.
DR	InterPro;	IPIR001791; Laminin_G.
DR	InterPro;	IPIR003129; TSPN.
DR	Pfam;	PF01391; Collagen; 7.
DR	Pfam;	PF02210; TSPN; 1.
DR	ProDom;	PD000007; Ctg_helix; 1.
DR	SMART;	SMO0282; Lamep; 1.
DR	SMART;	SMO0210; TSPN; 1.
KW	Extracellular matrix;	Connective tissue; Repeat; Hydroxylation;
KW	Cell adhesion;	Collagen; Glycoprotein; Signal; Alternative splicing;
KW	Polymorphism;	3D-structure.
FT	SIGNAL	1..23 POTENTIAL.
FT	CHAIN	24..1516 COLLAGEN ALPHA 1(XVII) CHAIN.
FT	CHAIN	1334..1516 ENDOSPATIN.
FT	DOMAIN	24..516 NONHELICAL REGION 1 (NC1).
FT	DOMAIN	517..550 TRIPLE-HELICAL REGION 1 (COL1).
FT	DOMAIN	551..560 NONHELICAL REGION 2 (NC2).
FT	DOMAIN	561..640 TRIPLE-HELICAL REGION 2 (COL2).
FT	DOMAIN	641..664 NONHELICAL REGION 3 (NC3).
FT	DOMAIN	665..786 TRIPLE-HELICAL REGION 3 (COL3).
FT	DOMAIN	787..809 NONHELICAL REGION 4 (NC4).
FT	DOMAIN	810..892 TRIPLE-HELICAL REGION 4 (COL4).
FT	DOMAIN	893..906 NONHELICAL REGION 5 (NC5).
FT	DOMAIN	907..948 TRIPLE-HELICAL REGION 5 (COL5).
FT	DOMAIN	949..961 NONHELICAL REGION 6 (NC6).

```
FT DOMAIN 962 1034 TRIPLE-HELICAL REGION 6 (COL6).
FT DOMAIN 1035 1044 NONHELICAL REGION 7 (NC7).
FT DOMAIN 1045 1077 TRIPLE-HELICAL REGION 7 (COL7).
FT DOMAIN 1078 1089 NONHELICAL REGION 8 (NC8).
FT DOMAIN 1090 1111 TRIPLE-HELICAL REGION 8 (COL8).
FT DOMAIN 1112 1118 NONHELICAL REGION 9 (NC9).
FT DOMAIN 1119 1173 TRIPLE-HELICAL REGION 9 (COL9).
FT DOMAIN 1174 1186 NONHELICAL REGION 10 (NC10).
FT DOMAIN 1187 1204 TRIPLE-HELICAL REGION 10 (COL10).
FT DOMAIN 1205 1516 NONHELICAL REGION 11 (NC11).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 691 691 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1329 1329 O-LINKED (GALNAC. . .) (POTENTIAL).
FT DISULFID 1366 1506 BY SIMILARITY.
FT DISULFID 1468 1498 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1095 1097 MISSING (in Isoform Short).
FT VARSPIC 1 180 /FTID-VSP_001155.
FT VARSPIC 181 215 /FTID-VSP_001156.
FT VARSPIC 181 215 /FTID-VSP_001156.
FT VARIANT 1437 1437 D -> N (increased risk of developing
FT VARIANT 1437 1437 prostate cancer).
FT VARIANT 1437 1437 /FTID-VAR_012709.
FT CONFLICT 428 428 F -> S (IN REF. 2).
FT CONFLICT 841 841 I -> V (IN REF. 2).
FT CONFLICT 877 877 V -> L (IN REF. 3).
FT CONFLICT 886 886 P -> R (IN REF. 3).
FT CONFLICT 912 912 P -> R (IN REF. 3).
FT CONFLICT 933 933 R -> L (IN REF. 3).
FT CONFLICT 975 975 P -> L (IN REF. 3).
FT CONFLICT 1064 1064 A -> P (IN REF. 3).
FT CONFLICT 1084 1084 L -> K (IN REF. 3).
FT CONFLICT 1120 1120 P -> A (IN REF. 3).
FT CONFLICT 1123 1123 P -> A (IN REF. 3).
FT CONFLICT 1126 1126 P -> A (IN REF. 3).
FT CONFLICT 1206 1206 P -> PPGP (IN REF. 2).
FT CONFLICT 1304 1304 G -> GQ (IN REF. 3).
FT CONFLICT 1314 1314 R -> G (IN REF. 3).
FT CONFLICT 1323 1324 A -> G (IN REF. 3).
FT CONFLICT 1443 1443 LR -> CG (IN REF. 3).
FT CONFLICT 1483 1483 R -> T (IN REF. 4).
FT CONFLICT 1483 1483 S -> Y (IN REF. 4).
SQ SEQUENCE 1516 AA; 153840 MM; 3C70E29A4476EE76 CRC64;
```

Query Match 8.4%; Score 102; DB 1; Length 1516;
Best Local Similarity 27.3%; Pred. No. 22;
Matches 47; Conservative 9; Mismatches 58; Indels 58; Gaps 9;

77 IQASMAPAAALPPVSPPAAPSVADNL---LASSDAALASWASL--LEDLSHTEGLSQ 130
Db 660 LRLALNGP-RGPPGPPGPGVPGIPGPGFRFGVNSSDVPGAGIPGVPGREGPPGPPGLPG 718
QY 131 APOPLADEGPPGR---SIG--GAAPSGALDLGLPATGCLLDGLGLEDIDTSMTD 183
Db 719 PPGPPGREGPPGPPRTGKGSLGEGAGPHGSGKAGPAP-GARGESGLAG----- 766
QY 184 NEIMAPASBGLKPGPEBDGPKGEAPLDEALDYLMVDLVGTQALERRPPPG 235
Db 767 ---APGPAG-PPGPPGPPG-----PPGP 786

Search completed: August 19, 2003, 09:33:10
Job time : 15 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:31:59 ; Search time 97 Seconds
(without alignments)
627,839 Million cell updates/sec

Title: US-09-701-675A-3

Perfect score: 1210
Sequence: 1 MSLGKLRKREERKEPLA.....YLMVLYGTQALRPPGPR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTRMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1210	100.0	236	4 Q9UHV2	Q9UHV2 homo sapien
2	1206	99.7	236	4 Q9BUE7	Q9BUE7 homo sapien
3	1044	86.3	236	11 Q9ULI0	Q9ULI0 mus musculu
4	1032	85.3	236	11 Q9D888	Q9D888 mus musculu
5	968	80.0	223	11 Q925E6	Q925E6 mus musculu
6	792	65.5	244	11 Q9DC22	Q9DC22 mus musculu
7	215	17.8	309	11 Q91WL3	Q91WL3 mus musculu
8	215	17.8	309	11 Q91JG5	Q91JG5 mus musculu
9	215	17.8	309	11 Q9C609	Q9C609 mus musculu
10	214	17.7	197	11 Q9ERC3	Q9ERC3 mus musculu
11	213	17.6	196	4 Q96CQ2	Q96CQ2 homo sapien
12	213	17.6	196	4 Q9UW9	Q9UW9 homo sapien
13	170	14.0	241	4 Q9TB18	Q9TB18 homo sapien
14	165	13.6	241	4 Q9BXL8	Q9BXL8 homo sapien
15	152	12.6	237	11 Q9CWM2	Q9CWM2 mus musculu
16	134	11.1	237	4 Q9NMK7	Q9NMK7 homo sapien

17	125.5	10.4	181	11 Q925E5	Q925E5 mus musculu
18	121.5	10.0	1339	11 Q35788	Q35788 rattus norv
19	120.5	10.0	752	10 Q8LK21	Q8LK21 chlamydomon
20	120	9.9	3164	12 Q69088	Q69088 human herpe
21	117	9.7	446	11 Q8YD23	Q8YD23 mus musculu
22	115.5	9.5	840	4 Q8N3F8	Q8N3F8 homo sapien
23	115	9.5	698	10 Q9FE49	Q9FE49 chlamydomon
24	115	9.5	699	10 Q9FED4	Q9FED4 chlamydomon
25	114	9.4	336	4 Q8IXT4	Q8IXT4 homo sapien
26	114	9.4	1313	5 Q9NE28	Q9NE28 leishmania
27	114	9.4	2137	4 Q15021	Q15021 homo sapien
28	113	9.3	983	11 Q9QXW1	Q9QXW1 mus musculu
29	112	9.3	715	4 Q9H0L5	Q9H0L5 homo sapien
30	112	9.3	791	4 Q9BY92	Q9BY92 homo sapien
31	110	9.1	202	11 Q921E8	Q921E8 homo sapien
32	109	9.0	657	4 Q8N3N8	Q8N3N8 homo sapien
33	109	9.0	688	12 Q65575	Q65575 bovine herp
34	108.5	9.0	1531	4 Q60336	Q60336 bovine herp
35	108	8.9	941	16 Q9Z1W5	Q9Z1W5 streptomyce
36	107	8.8	993	11 Q8CHE2	Q8CHE2 mus musculu
37	107	8.8	1317	11 Q35211	Q35211 mus musculu
38	107	8.8	1317	11 Q8CB63	Q8CB63 mus musculu
39	106.5	8.8	512	17 Q9HPL5	Q9HPL5 halobacteri
40	106.5	8.8	753	16 Q87124	Q87124 pseudomonas
41	106.5	8.8	1245	4 Q43636	Q43636 homo sapien
42	106.5	8.8	1251	4 Q9UMG2	Q9UMG2 homo sapien
43	106	8.8	442	9 Q8HAP5	Q8HAP5 burkholderi
44	106	8.8	993	4 Q9HC9	Q9HC9 homo sapien
45	104	8.6	539	11 Q8CFW3	Q8CFW3 mus musculu

ALIGNMENTS

RESULT 1
Q9UHV2 PRELIMINARY; PRT; 236 AA.
ID Q9UHV2
AC Q9UHV2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE CDK4-binding protein p34SEI1 (TRIP-Brl) (SCDK4-binding protein p34SEI1).
GN SEI1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=20047903; PubMed=10580009;
RA Sugimoto M., Nakamura T., Ohtani N., Hampson I.N.,
RA Shimamoto A., Furutachi Y., Okumura K., Niva S., Taya Y., Hara E.,
RT "Regulation of CDK4 activity by a novel CDK4 binding protein, p34SEI1.";
RT Genes Dev. 13:3027-3033(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21231173; PubMed=11331592;
RA Hsu S.T., Yang C.M., Sim K.G., Hentschel D.M., O'Leary E.,
RA Bonventre J.V.,
RT "TRIP-Brl: a novel family of PHD zinc finger- and bromodomain-interacting proteins that regulate the transcriptional activity of ERF-1/DP-1.";
RT EMBO J. 20:2273-2285(2001).
RN [3]
RP SEQUENCE FROM N.A.
RA Rieder M.J., Braun A.C., Montoya M.A., Chung M.-W., Nguyen C.P.,
RA Nguyen D.A., Livingston R.J., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.,
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF117959; AAF08349.1; -;
EMBL; AF366402; AAK52831.1; -;

```

DR EMBL: AY130860.1; -.
SQ SEQUENCE 236 AA; 24673 MW; 92F25EA328F155B1 CRC64;

Query Match 100.0%; Score 1210; DB 4; Length 236;
Best Local Similarity 100.0%; Pred. No. 7.9e-86;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTLKIHHSLOOS 60
DB 1 MLSKGLKRRKEEKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTLKIHHSLOOS 60
QY 61 EDDLRLHVLVYNTLRRIOASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
DB 61 EDDLRLHVLVYNTLRRIOASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
QY 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
DB 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
QY 181 MYDNELMAPASEGLKPGPEDGPGKEAPELDEAELDYLMYLVGTQALERPPGGR 236
DB 181 MYDNELMAPASEGLKPGPEDGPGKEAPELDEAELDYLMYLVGTQALERPPGGR 236

RESULT 3
Q9JUL10 PRELIMINARY; PRT; 236 AA.
ID 09JUL10
AC 09JUL10
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DR 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
P3ASE1-1 (1110032C13RIK protein) (TRIP-BR1) (RIKEN CDNA 1110032C13

```

```

DE gene).
GN SE11 OR 1110032C13RIK.
OS Mus musculus (Mouse).
OC Mammalia; Eutelestoma; Chordata; Craniata; Vertebrata; Eutelestoma;
OC Mammalia; Eutelestoma; Chordata; Craniata; Vertebrata; Eutelestoma;
OX NCBI_Taxid=10090;

[1]
RP SEQUENCE FROM N.A.
RT Ohtani N., Hara E.;
RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shigaoka A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stanbly F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlay M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombert P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).

[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21231173; PubMed=11331592;
RA Hsu S.I., Yang C.M., Sim K.G., Hentschel D.M., O'Leary E.,
RA Bonventre J.V.;
RT "TRIP-BR: a novel family of PHD zinc finger- and bromodomain-
RT interacting proteins that regulate the transcriptional activity of
RT E2F-1/DP-1.";
RL EMBL J. 20:2273-2285(2001).

[4]
RP SEQUENCE FROM N.A.
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
RA Strausberg R.;
RX EMBL: AF218291; AAF37653.1; -.
DR EMBL: AK004022; BAB23130.1; -.
DR EMBL: AF366401; AAK52830.1; -.
DR EMBL: BC016077; AAH16077.1; -.
DR MGI:1913438; 1110032C13RIK.
SQ SEQUENCE 236 AA; 25136 MW; 989ADF829D84C5 CRC64;

Query Match 86.3%; Score 1044; DB 11; Length 236;
Best Local Similarity 86.4%; Pred. No. 5.2e-73;
Matches 204; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

QY 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTLKIHHSLOOS 60
DB 1 MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAAPPVASSSLFDLSVTLKIHHSLOOS 60
QY 61 EDDLRLHVLVYNTLRRIOASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
DB 61 EDDLRLHVLVYNTLRRIOASMAPAALPPVSPPAASVADNLLASSDAALSASMASLLE 120
QY 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
DB 121 DLSHTEGISOAPQPLADGPPGRSISGAAPSLGALDILGPATGCLLDGLEGLEFEDIDTS 180
QY 181 MYDNELMAPASEGLKPGPEDGPGKEAPELDEAELDYLMYLVGTQALERPPGGR 236
DB 181 MYDNELMAPASEGLKPGPEDGPGKEAPELDEAELDYLMYLVGTQALERPPGGR 236

```

DB 181 MYDSELMLPASEGLKPGENGPAKKEPPELDEAELDYMDVLYGTQALERPPEGGR 236

RESULT 4
Q9D888 PRELIMINARY; PRT; 236 AA.

AC Q9D888; 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DE 1110032C13Rik protein.
GN 1110032C13Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57Bl/6J; TISSUE=Small Intestine;
RX MEDLINE=21085660; PubMed=11217851;
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Glassi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyoko-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK008303; BAB25588.1;
DR MGD: MGI:1913438; 1110032C13Rik.
SQ SEQUENCE 236 AA; 25023 MW; 95AE0BCFAC477549 CRC64;

Query Match 85.3%; Score 1032; DB 11; Length 236;
Best Local Similarity 85.6%; Pred. No. 4.4e-72;
Matches 202; Conservative 10; Mismatches 24; Indels 0; Gaps 0;

QY 1 MTSKGLKRRREEREEKPLAVDSWMDPGHAAYQAQPPAVASSSLFDLSVYKLHSLQOS 60
DB 1 MTSKGLKRRREEREEKPLAVDSWMDPGHAAYQAQPPAVASSSLFDLSVYKLHSLQOS 60
QY 61 EPDLRLVLYVNTLRRIQASMAAPALPPVSPAPASVADNLASSDAALSASMSLLE 120
DB 61 EPDLRLVLYVNTLRRIQASMAAPALPPVSPAPASVADNLASSDAALSASMSLLE 120
QY 121 DLSHIEGLSQADPLADESGPPGRSIGGAAPSLGALDLPATGCLLDGEGLEFEDIDTS 180
DB 121 DLSHIEGLSQADPLADESGPPGRSIGGAAPSLGALDLPATGCLLDGEGLEFEDIDTS 180
QY 181 MYDNEIMAPASGLKPGPDGCKEAPPELDEAELDYMDVLYGTQALERPPEGGR 236
DB 181 MYDNEIMAPASGLKPGPDGCKEAPPELDEAELDYMDVLYGTQALERPPEGGR 236
QY 181 MYDSELMLPASEGLKPGENGPAKKEPPELDEAELDYMDVLYGTQALERPPEGGR 236
DB 181 MYDSELMLPASEGLKPGENGPAKKEPPELDEAELDYMDVLYGTQALERPPEGGR 236

RESULT 5
Q925E6 PRELIMINARY; PRT; 223 AA.

AC Q925E6; 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
TRIP-Bit (Fragment).

GN 1110032C13Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57Bl/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851;
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Glassi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyoko-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

Query Match 80.0%; Score 968; DB 11; Length 223;
Best Local Similarity 85.2%; Pred. No. 3.6e-67;
Matches 190; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

QY 1 MTSKGLKRRREEREEKPLAVDSWMDPGHAAYQAQPPAVASSSLFDLSVYKLHSLQOS 60
DB 1 MTSKGLKRRREEREEKPLAVDSWMDPGHAAYQAQPPAVASSSLFDLSVYKLHSLQOS 60
QY 61 EPDLRLVLYVNTLRRIQASMAAPALPPVSPAPASVADNLASSDAALSASMSLLE 120
DB 61 EPDLRLVLYVNTLRRIQASMAAPALPPVSPAPASVADNLASSDAALSASMSLLE 120
QY 121 DLSHIEGLSQADPLADESGPPGRSIGGAAPSLGALDLPATGCLLDGEGLEFEDIDTS 180
DB 121 DLSHIEGLSQADPLADESGPPGRSIGGAAPSLGALDLPATGCLLDGEGLEFEDIDTS 180
QY 181 MYDNEIMAPASGLKPGPDGCKEAPPELDEAELDYMDVLY 223
DB 181 MYDSELMLPASEGLKPGENGPAKKEPPELDEAELDYMDVLY 223

RESULT 6
Q9DC22 PRELIMINARY; PRT; 244 AA.

AC Q9DC22; 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE 1110032C13Rik protein.
GN 1110032C13Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57Bl/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851;
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Glassi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyoko-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wyszewski A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection,"
 RL Nature 409:685-690(2001).
 DR EMBL: AK002324; BAB2201.1;
 DR MGD: MGI:1913438; 1110032C13R1K.
 SQ SEQUENCE 244 AA; 26068 MW; 0F464F6419D362A6 CRC64;

Query Match 65.5%; Score 792; DB 11; Length 244;
 Best Local Similarity 71.7%; Pred. No. 1.6e-53;
 Matches 170; Conservative 10; Mismatches 55; Indels 2; Gaps 2;

OY 1 MLSKGLKRRKEEKEPELAVDSWMLDPGHAVAQAAPVAVSSSLFEDSVLKHLHNS 60
 DB 1 MLSKGLKRRKEEKEPELAVDSWMLDPGHAVAQAAPVAVSSSLFEDSVLKHLHNS 60
 OY 61 EPDLRLHLVYVNTLRIRIOASMAAPALPPVSPAPVADNLASSDAALASMASLILE 120
 DB 61 EPDLRLHLVYVNTLRIRIOASMAAPALPPVSPAPVADNLASSDAALASMASLILE 120
 OY 121 DLSHTEGTSQAQPLADGPPGKSTIGAAPSLGALDILG-PATGCLLDGLEGLEFEDIDT 179
 DB 121 DLSHTEGTSQAQPLADGPPGKSTIGAAPSLGALDILG-PATGCLLDGLEGLEFEDIDT 179
 OY 121 DLSHTEGTSQAQPLADGPPGKSTIGAAPSLGALDILG-PATGCLLDGLEGLEFEDIDT 180
 DB 121 DLSHTEGTSQAQPLADGPPGKSTIGAAPSLGALDILG-PATGCLLDGLEGLEFEDIDT 180
 OY 180 SMYDNEMLAPASEGLKPGEDGPGKEAPELDEALDYLMDVYVTOALERPGRGR 236
 DB 181 CTYVNYGYQPLRVS-SAAPENGPAKEPELDEALDYLMDVYVTOALERPGRGR 236

RESULT 7
 O91WL3 PRELIMINARY; PRT; 309 AA.

AC O91WL3:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 33.3 kDa protein.
 GN AB041541.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Strausberg R.;
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC014726; AAH14726.1;
 DR MGD: MGI:1931026; AB041541.
 KW Hypothetical protein.
 SQ SEQUENCE 309 AA; 33328 MW; ED3B5B38AC839D5 CRC64;

Query Match 17.8%; Score 215; DB 11; Length 309;
 Best Local Similarity 27.8%; Pred. No. 7.9e-09;
 Matches 90; Conservative 33; Mismatches 87; Indels 114; Gaps 12;

OY 1 MLSGGLKRRKEEKEPELAVDSWMLDPGHAVAQAAPVAVSSSLFEDSVLKHLHNS 56
 DB 1 MLSGGLKRRKEEKEPELAVDSWMLDPGHAVAQAAPVAVSSSLFEDSVLKHLHNS 51
 OY 57 LQSEPDRLHLVYVNTLRIRIOASMAAPALPPVSPAPVADNLASSDAALASMASLILE 93
 DB 57 LQSEPDRLHLVYVNTLRIRIOASMAAPALPPVSPAPVADNLASSDAALASMASLILE 93
 OY 52 RPLTEPSLQKTVLNNMLRQIEELKQBGSLRPAFTSSOPNSLSYQAPPAAPHC 111
 DB 52 RPLTEPSLQKTVLNNMLRQIEELKQBGSLRPAFTSSOPNSLSYQAPPAAPHC 111
 OY 94 -----PAASVADNLASSDAALASMASLILE 120
 DB 112 DLGSTPLEACTLPASLLEDDNDPCTQLQSVHPAAPT-----RLSSAALPAKDSFSS 165
 OY 121 DLSHTEGTSQAQPLADGPPGKSTIGAAPSLGALDILG-PATGCLLDGLEGLEFEDIDT 157
 DB 121 DLSHTEGTSQAQPLADGPPGKSTIGAAPSLGALDILG-PATGCLLDGLEGLEFEDIDT 157
 OY 166 ALDIEELCPTSTSTEAHTAAPGPKGTSSSESVORPEGEERTDSDRPMDSLPGNFE 225
 DB 166 ALDIEELCPTSTSTEAHTAAPGPKGTSSSESVORPEGEERTDSDRPMDSLPGNFE 225
 OY 158 LGPATGCLLDGLEGLEFEDIDTSMYDNEMLAPAS-----EGLKPPEDGPGKEAP----- 208
 DB 158 LGPATGCLLDGLEGLEFEDIDTSMYDNEMLAPAS-----EGLKPPEDGPGKEAP----- 208

DB 226 ITTSTGFLTDLTLDLLEFADIDTSMYDNPCTSSASGASKAPVAVSADLTKTAPYSNOP 285
 OY 209 -----BLDEAEIDYLDVYVGT 225
 DB 286 VAPSQPFKMDLTDLHINEVLYGS 309

RESULT 8
 O9JUG5 PRELIMINARY; PRT; 309 AA.

AC O9JUG5:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Brain cDNA, clone MNCB-1504, similar to D50917 KIAA0127 protein (Homo sapiens).
 GN AB041541.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL; TISSUE=Brain;
 RA Osada N., Kusuda Y., Tanuma R., Ito A., Hirata M., Sugano S.,
 RA Hashimoto K.;
 RT "Isolation of full-length cDNA clones from mouse brain cDNA library made by oligo-capping method."
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB041541; BA95026.1;
 DR MGD: MGI:1931026; AB041541.
 SQ SEQUENCE 309 AA; 33282 MW; D5A78698E1DE9F11 CRC64;

Query Match 17.8%; Score 215; DB 11; Length 309;
 Best Local Similarity 27.8%; Pred. No. 7.9e-09;
 Matches 90; Conservative 33; Mismatches 87; Indels 114; Gaps 12;

OY 1 MLSGGLKRRKEEKEPELAVDSWMLDPGHAVAQAAPVAVSSSLFEDSVLKHLHNS 56
 DB 1 MLSGGLKRRKEEKEPELAVDSWMLDPGHAVAQAAPVAVSSSLFEDSVLKHLHNS 51
 OY 57 LQSEPDRLHLVYVNTLRIRIOASMAAPALPPVSPAPVADNLASSDAALASMASLILE 93
 DB 57 LQSEPDRLHLVYVNTLRIRIOASMAAPALPPVSPAPVADNLASSDAALASMASLILE 93
 OY 52 RPLTEPSLQKTVLNNMLRQIEELKQBGSLRPAFTSSOPNSLSYQAPPAAPHC 111
 DB 52 RPLTEPSLQKTVLNNMLRQIEELKQBGSLRPAFTSSOPNSLSYQAPPAAPHC 111
 OY 94 -----PAASVADNLASSDAALASMASLILE 120
 DB 112 DLGSTPLEACTLPASLLEDDNDPCTQLQSVHPAAPT-----RLSSAALPAKDSFSS 165
 OY 121 DLSHTEGTSQAQPLADGPPGKSTIGAAPSLGALDILG-PATGCLLDGLEGLEFEDIDT 157
 DB 121 DLSHTEGTSQAQPLADGPPGKSTIGAAPSLGALDILG-PATGCLLDGLEGLEFEDIDT 157
 OY 166 ALDIEELCPTSTSTEAHTAAPGPKGTSSSESVORPEGEERTDSDRPMDSLPGNFE 225
 DB 166 ALDIEELCPTSTSTEAHTAAPGPKGTSSSESVORPEGEERTDSDRPMDSLPGNFE 225
 OY 226 ITTSTGFLTDLTLDLLEFADIDTSMYDNPCTSSASGASKAPVAVSADLTKTAPYSNOP 285
 DB 226 ITTSTGFLTDLTLDLLEFADIDTSMYDNPCTSSASGASKAPVAVSADLTKTAPYSNOP 285
 OY 209 -----ELDEAEIDYLDVYVGT 225
 DB 286 VAPSQPFKMDLTDLHINEVLYGS 309

RESULT 9

O8C609 PRELIMINARY; PRT; 309 AA.

AC O8C609:
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 ON NCBI_TaxId=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=2354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL: AK076787; BAC36480.1; -
 KM Hypothetical protein.
 SQ SEQUENCE 309 AA; 33312 MW; D4178688F0DF8F00 CRC64;
 Query Match 17.8%; Score 215; DB 11; Length 309;
 Best Local Similarity 27.8%; Pred. No. 7, 9e-09;
 Matches 90; Conservative 33; Mismatches 87; Indels 114; Gaps 12;
 QY 1 MUKGKLRREEREEKEPLAVDSWMLDPGHAAPVAVAS--SSLFDSLVTKLHNS 56
 DB 1 MUKGKGRKRFDEHED-----GLEGKIVSPDQSPKSVYTLQRTIFNISTLMKLYNH 51
 QY 57 LQOSEPDRLHVLVYVNTLRRIQ-----ASMAPA-----AALPPVSP- 93
 DB 52 RPLTEPSLQKTVLINMLRRIQEELKQEGSLRPAFTSSQPSNSLSYQEAAPPAPHPHC 111
 QY 94 -----PAAPSYADNLLASDAALASMSALILE 120
 DB 112 DGGSTTPLEACLTPLASLEDDNDTRCTIQAVHPAAPT-----RUSALALREKDSFSS 165
 QY 121 DLSHTEGL-----SOAQPPLADEGPPGRSIGA-----APSLGALD-----L 157
 DB 166 ALDEIEELCPTSTSTPAHTAPEGPKGTSSSSVQKPGEPGRTDGRFMDSLPGNFE 225
 QY 158 LGPANGCCLLDGLEG-LFEDIDTSMYDNLNAPAS---EGKAPGPDGCKEAP----- 208
 DB 226 ITTSTGFTLDTLDFADIDTSMYDFDPCTSASGASKMAPVSADLLKTLAPYSNOP 285
 QY 209 -----ELDEAELDYIMDYLVGT 225
 DB 286 VAPSQPFKMDLTRELDHIMEVLVGS 309
 RESULT 10
 Q9ERC3 PRELIMINARY; PRT; 197 AA.
 AC Q9ERC3;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE Replication protein-binding trans-activator RBT1.
 DE Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxId=10090;
 RP SEQUENCE FROM N.A.
 RC Cho J.M., Alaoui-Jamali M.A.;
 RT "Cloning of mouse RBT1 cDNA."
 RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Straussberg R.;
 RA Submitted (Jul-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AF317202; AG30951.1; -
 DR EMBL: BC034886; AAH34886.1; -
 DR InterPro: IPR000194; ATPase_a/centr.
 DR PROSITE: PS00152; ATPase_ALPHA_BETA; 1.
 SQ SEQUENCE 197 AA; 21934 MW; CFF6FAAC0A91D7E7 CRC64;
 Query Match 17.7%; Score 214; DB 11; Length 197;

Best Local Similarity 31.8%; Pred. No. 5, 5e-09;
 Matches 74; Conservative 37; Mismatches 72; Indels 50; Gaps 13;
 QY 5 GLKRRKREEREEKEPLAVDSW-MLDPGHAAPVAVAS--SSLFDSLVTKLHNSIQ 58
 DB 3 GLKRRKSDLEEEEEE-----KWDW-----SPYLRSTQOALLRISLTKVRSIG 47
 QY 59 QSEPDRLHVLVYVNTLRRIQAS--MAPAALPPVSPPAAPSVADNLLASDAALASMA 116
 DB 48 PRAPSLRRVHLHNTLQQLQAALRLAPAPALPPEPL-----FLGEEFSLSTTIG 97
 QY 117 SLLEDL-SHTEGLSQAPPLADEGPPGRSIGAAPSLGALLDLPATGCLLDGLELFE 175
 DB 98 SLIRELDSMDEEPPPLNPAASSPONEIVSQADPVF--LEAL--SSRYLGDPSGLDFFL 153
 QY 176 DIDTSMYDNELMAPASEGLKPGED--GPGKEAPELDEALDYIMDYLVGT 225
 DB 154 DIDTSAVERKQVALPPE-----PPHSIFCSGSGWE-----WNELDHIMEITLGS 197
 RESULT 11
 Q9ECQ2 PRELIMINARY; PRT; 196 AA.
 AC Q9ECQ2;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxId=9606;
 RP SEQUENCE FROM N.A.
 RC Straussberg R.;
 RA Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL: BC014061; AAH14061.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 196 AA; 21768 MW; 9C54AFA126F17AF1 CRC64;
 Query Match 17.6%; Score 213; DB 4; Length 196;
 Best Local Similarity 32.6%; Pred. No. 6, 5e-09;
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;
 QY 5 GLKRRKREEREEKEPLAVDSW-MLDPGHAAPVAVAS--SSLFDSLVTKLHNSIQSEPD 63
 DB 4 GLKRRKSDLEEEEEE-----ERWESPAGLSYQQA-----LRTSLDRVQNSLGRAPS 51
 QY 64 LRHLVLVYVNTLRRIQAS--MAPAALPPVSPPAAPSVADNLLASDAALASMSALIED 121
 DB 52 LRRLVHLHNTLQQLQAALRLAPAPALPPEPL-----FLGEEFSLSTTIGSLNE 101
 QY 122 L-SHTEGLSQAPPLADEGPPGRSIGAAPSLGALLDLPATGCLLDGLELFE 175
 DB 102 LDTSMYDNELMAPASEGLKPGED--GPGKEAPELDEALDYIMDYLVGT 225
 QY 176 DIDTSMYDNELMAPASEGLKPGED--GPGKEAPELDEALDYIMDYLVGT 225
 DB 153 DIDTSAVERKQVALPPE-----PPHSIFCSGSGWE-----WNELDHIMEITLGS 196
 RESULT 12
 Q9UW9 PRELIMINARY; PRT; 196 AA.
 AC Q9UW9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE RPA-binding trans-activator.
 GN RBT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20440390; PubMed=10982866;
 RA Cho J.M., Song D.J., Alouali-Jamali M.A.;
 RT "Rnt1, a novel transcriptional co-activator, binds the second subunit
 of Replication Protein A.";
 RL Nucleic Acids Res. 28:3478-3485(2000).
 DR EMBL; AF192529; AAF05761.1; -
 SQ SEQUENCE 196 AA; 21798 MW; 33433EF7F8A9EBA7 CRC64;

Query Match 17.6%; Score 213; DB 4; Length 196;
 Best Local Similarity 32.6%; Pred. No. 6.5e-09;

Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

OY 5 GLKRREREKEPEPLAVDSW-WIDPGHAAVAQAAPPVASSSLFDVSLKHLQSEPD 63
 DB 4 GLKRRHSDLEEE-ERMESPPAGLSYQQA-----LKRISLKVQKSLGPPRAPS 51
 OY 64 LRHLVLYVNTLRRIQAS--MAPAALPPVSPPPAASVADNLASSDAALSASMSLLED 121
 DB 52 LRHVLHNTLQQLQAALRLAPALPPEPL-----FLGDEPDSLSTIGSILRE 101
 OY 122 L-SHTEGSLQAPQ-----LADGPPGSGISGAASLADLGLGATGCLDDGLELFE 175
 DB 102 LDTSDGTEPPONPYTPVIGLQNEVP-----QDPVFLKLV--SSRYLGDSGLDDPEFL 152
 OY 176 DITSMYNEMLAPASEGLKPGPED---GPKGEAPELDEALDYLMDVLVGT 225
 DB 153 DITSAVVEKE---PARAPPEP-PHNLFCAPGSM-----WNELDHIMEIILIGS 196

RESULT 13

OQTB18 PRELIMINARY; PRT; 241 AA.

AC 08TB18; 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC025263; AAK25263.1; -
 DR Gene; HGNC:14625; CDCA4.
 KW Hypothetical protein.
 SQ SEQUENCE 241 AA; 26112 MW; CC936EDED4A9A17 CRC64;

Query Match 14.0%; Score 170; DB 4; Length 241;
 Best Local Similarity 28.0%; Pred. No. 1.7e-05;

Matches 74; Conservative 38; Mismatches 90; Indels 62; Gaps 13;

OY 1 MLKSGIKRR--REEEKEPLA---VDSMWLDPGHAAVAQAAPPVASSSLFDVSLKHL 54
 DB 1 MFARGLKRCYGHEDVSGALAGLKTSSY-----SLQKSLDMSLYVKLQ 46
 OY 55 HSLQOSEPDRLHLVYVNTLRRIQASMAPAALPPVSPPPAASVADNLASSDAALSAS 114
 DB 47 LCHMLVEPNLCRSVLIANTVRIQIEMTQDGTWRTV-APQAERAPLRLVSTETLCRAA 105
 OY 115 M-----ASLLEDLSHTEG-----LSQAQPLADG-----PPGRSISGAAPSLG- 153
 DB 106 WGQECAPHPGLGD-GHTQGPVSDICPVYSAQAPRHILQSSAMEMDGPENKGSFHKSLDQ 164
 OY 154 ---ALDLGPATGCLDDGLGLEDITDSMYD-NELWAPASEGLKPGPEGD----- 201

DB 165 IFETLETNPS--C-----MEELFSVDSPYYDLDITVLTGMGAPPGCGLEGAPAT 217
 OY 202 PKEKEAPELDEALDYLDMDVLVGT 225
 DB 218 PGPSSCKSDLGELDHVEILVET 241

RESULT 14

OQBYL8 PRELIMINARY; PRT; 241 AA.

AC 09BYL8; 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hematopoietic progenitor protein.
 GN HEPP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21375891; PubMed=11482882;
 RA Abdullah J.M., Jing X., Spassov D.S., Nachman R.G., Jurecic R.;
 RT "Cloning and Characterization of Hepp, a Novel Gene Expressed
 Preferentially in Hematopoietic Progenitors and Mature Blood Cells.";
 RL Blood Cells Mol. Dis. 27:667-676(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=B-cell;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF322239; AAK31075.1; -
 DR EMBL; BC011736; AAK11736.1; -
 SQ SEQUENCE 241 AA; 26113 MW; 9693647E77865D7A CRC64;

Query Match 13.6%; Score 165; DB 4; Length 241;
 Best Local Similarity 27.7%; Pred. No. 4.2e-05;

Matches 73; Conservative 39; Mismatches 90; Indels 62; Gaps 13;

OY 1 MLKSGIKRR--REEEKEPLA---VDSMWLDPGHAAVAQAAPPVASSSLFDVSLKHL 54
 DB 1 MFARGLKRCYGHEDVSGALAGLKTSSY-----SLQKSLDMSLYVKLQ 46
 OY 55 HSLQOSEPDRLHLVYVNTLRRIQASMAPAALPPVSPPPAASVADNLASSDAALSAS 114
 DB 47 LCHMLVEPNLCRSVLIANTVRIQIEMTQDGTWRTV-APQAERAPLRLVSTETLCRAA 105
 OY 115 M-----ASLLEDLSHTEG-----LSQAQPLADG-----PPGRSISGAAPSLG- 153
 DB 106 WGQECAPHPGLGD-GHTQGPVSDICPVYSAQAPRHILQSSAMEMDGPENKGSFHKSLDQ 164
 OY 154 ---ALDLGPATGCLDDGLGLEDITDSMYD-NELWAPASEGLKPGPEGD----- 201
 DB 165 IFETLETNPS--C-----MEELFSVDSPYYDLDITVLTGMGAPPGCGLEGAPAT 217
 OY 202 PKEKEAPELDEALDYLDMDVLVGT 225
 DB 218 PGPSSCKSDLGELDHVEILVET 241
 RESULT 15
 OQCNM2 PRELIMINARY; PRT; 237 AA.
 AC 09CNM2; 09JUN96;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE ES cells cDNA, RIKEN full-length enriched library, clone:2410018C03
 DE product:hematopoietic progenitor protein, full insert sequence
 DE (Hematopoietic progenitor protein).
 GN 2410018C03RIK OR HEPP.
 OS Mus musculus (Mouse).

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:32:03 ; Search time 29 Seconds
(without alignments)
344.323 Million cell updates/sec

Title: US-09-701-675A-3
Perfect score: 1210
Sequence: 1 MLSKGIKRRKEEKEPELA.....YIMDLVVGQALERPDPGR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 s
```

```
1: Issue_Parents_A: *
2: /cgn2_6/prodata1/1aa/5A.COMB .pep: *
3: /cgn2_6/prodata1/1aa/5B.COMB .pep: *
4: /cgn2_6/prodata1/1aa/6A.COMB .pep: *
5: /cgn2_6/prodata1/1aa/6B.COMB .pep: *
6: /cgn2_6/prodata1/1aa/PCURS.COMB .pep: *
7: /cgn2_6/prodata1/1aa/backfiles1.pep: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	113	9.3	983	4	US-09-311-236-2	Sequence 2, Appl1
2	107	8.8	1317	3	US-09-083-521-7	Sequence 7, Appl1
3	106.5	8.8	760	4	US-09-252-991A-27790	Sequence 27790, Appl1
4	104	8.6	661	2	US-08-795-868-14	Sequence 14, Appl1
5	104	8.6	661	4	US-09-303-069-14	Sequence 14, Appl1
6	104	8.6	661	4	US-09-134-250-14	Sequence 14, Appl1
7	103.5	8.6	7257	3	US-09-335-409-5	Sequence 14, Appl1
8	103.5	8.6	7257	3	US-09-358-102-5	Sequence 5, Appl1
9	103.5	8.6	7257	4	US-09-567-969-5	Sequence 5, Appl1
10	103.5	8.6	7257	4	US-09-568-480-5	Sequence 5, Appl1
11	103.5	8.6	7257	4	US-09-568-482-5	Sequence 5, Appl1
12	103.5	8.6	7257	4	US-09-568-476-5	Sequence 5, Appl1
13	103.5	8.6	7257	4	US-09-567-899-5	Sequence 5, Appl1
14	100	8.3	655	3	US-08-857-076-57	Sequence 5, Appl1
15	100	8.3	1601	4	US-09-345-473E-40	Sequence 40, Appl1
16	99.5	8.2	476	3	US-09-189-035-1	Sequence 1, Appl1
17	99.5	8.2	476	3	US-09-382-086-1	Sequence 1, Appl1
18	98	8.1	417	4	US-09-559-360-88	Sequence 88, Appl1
19	97.5	8.1	830	4	US-09-562-737-37	Sequence 37, Appl1
20	96.5	8.0	1384	3	US-08-976-255-11	Sequence 11, Appl1
21	96	7.9	1298	2	US-08-650-874-2	Sequence 2, Appl1
22	96	7.9	1298	3	US-09-259-821A-2	Sequence 2, Appl1
23	96	7.9	1298	3	US-08-843-659-2	Sequence 2, Appl1
24	95.5	7.9	1257	1	US-08-340-428B-49	Sequence 49, Appl1
25	95	7.9	982	4	US-09-556-877-176	Sequence 176, Appl1
26	95	7.9	982	4	US-09-620-412C-176	Sequence 176, Appl1
27	95	7.9	982	4	US-09-558-419-176	Sequence 176, Appl1

28	95	7.9	1.006	4	US-09-556-877-190	Sequence 190, Appl
29	95	7.9	1.006	4	US-09-620-412C-190	Sequence 190, Appl
30	95	7.9	1.006	4	US-09-598-419-190	Sequence 190, Appl
31	94.5	7.8	436	4	US-09-252-991A-26982	Sequence 26982, Appl
32	94.5	7.8	830	4	US-09-562-737-31	Sequence 31, Appl
33	93.5	7.7	2972	4	US-09-576-181-2	Sequence 1, Appl
34	93.5	7.7	3118	4	US-09-579-181-1	Sequence 1, Appl
35	93	7.7	1135	4	US-08-094-948A-29	Sequence 29, Appl
36	93	7.7	1135	5	PCT-US96-09319-29	Sequence 29, Appl
37	93	7.7	1706	2	US-08-459-568-2	Sequence 2, Appl
38	93	7.7	1706	2	US-08-399-411-2	Sequence 2, Appl
39	93	7.7	1706	3	US-08-516-859A-2	Sequence 2, Appl
40	93	7.7	1706	3	US-09-586-472-2	Sequence 2, Appl
41	93	7.7	1706	4	US-09-528-706-2	Sequence 2, Appl
42	92	7.6	415	4	US-09-252-991A-1348	Sequence 1348, Appl
43	91.5	7.6	486	2	US-08-642-423-2	Sequence 2, Appl
44	91.5	7.6	486	4	US-08-930-915A-26	Sequence 26, Appl
45	91	7.5	1557	4	US-09-410-531B-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
 US-09-311-236-2
 : Sequence 2, Application US/09311236
 : Patent No. 6376215
 : GENERAL INFORMATION:
 : APPLICANT: Lathangue, Nicholas B
 : APPLICANT: Shikama, No. 63762151ko
 : TITLE OF INVENTION: The University Court of the University of Glasgow
 : TITLE OF INVENTION: JMY, a co-activator for p300/CBP, nucleic acid encoding
 : FILE REFERENCE: JMY and uses thereof
 : FILE REFERENCE: AHB/LP5/31300
 : CURRENT APPLICATION NUMBER: US/09/311,236
 : CURRENT FILING DATE: 1999-05-13
 : EARLIER APPLICATION NUMBER: GB 9722238.4
 : EARLIER FILING DATE: 1997-10-21
 : EARLIER APPLICATION NUMBER: GB 9618235.5
 : EARLIER FILING DATE: 1998-08-20
 : NUMBER OF SEQ ID NOS: 3
 : SOFTWARE: Patentin Ver. 2.0
 : SEQ ID NO 2
 : LENGTH: 983
 : TYPE: PRT
 : ORGANISM: Mus sp.
 : US-09-311-236-2

Query Match	9.3%;	Score 113;	DB 4;	Length 983;
Best Local Similarity	25.1%;	Pred. No. 0.044;		
Matches	61;	Conservative	26;	Mismatches 84;
			Indels	72;
			Gaps	11

[illegible]

```
RESULT 2
US-09-083-521-7
; Sequence 7, Application US/09083521
; Patent No. 6048970
; GENERAL INFORMATION:
; APPLICANT: IAL, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,521
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0527 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 853-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1317 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 2459993
; US-09-083-521-7

Query Match      8.8%; Score 107; DB 3; Length 1317;
Best Local Similarity 23.7%; Pred. No. 0.24;
Matches 58; Conservative 20; Mismatches 79; Indels 88; Gaps 10;

OY 78 QASMAPAALPPVSP-----PAPSVDN-----LLASDA 109
DB 746 EGSAPPOLPLPSVSPSCGASLDSEASAPDILPASPTPAAGSWTVPEPAPTLESSGS 805
OY 110 ALSASMASLIEDLSHTEGLSQAPQPLADEGPPGRSIG-----GAAPSIGALDL 157
DB 806 SLGGEAPB-SEDEPTTEATSGVFDLSSDGRHTEKSGIVPALRSIQKVGCFRDLSDSLDI 864
OY 158 LGPAT--GC-----LDDGL-----ELFEDIDTSMNDNELMAPAS 191
DB 865 PSSASDGCCEVLSPSAGPPGCGPRAVDGYDTENYESPEFVLKHAHSESEPEAFGEFAS 924
OY 192 EGLKPGPD-----GPKKEAPDELDEA---ELDYIMDYLV-----TQALR 230
DB 925 EGESBGPPLLSVLSGLSKSPYRDSAYFSDLDASEPTTGPPEKHSGIDDSQKEODLRS 984
OY 231 PPGPG 235
DB 985 PPSPG 989
```

```
RESULT 3
US-09-252-991A-27790
; Sequence 27790, Application US/09252991A
; Patent No. 6551795
```

```
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUDINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27790
; TYPE: PRT
; LENGTH: 760
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-27790

Query Match      8.8%; Score 106.5; DB 4; Length 760;
Best Local Similarity 25.7%; Pred. No. 0.13;
Matches 59; Conservative 29; Mismatches 101; Indels 41; Gaps 10;

OY 2 LSKGLKRRKEE--EKEEPLAVDSWMLDPGH-AAVAQAPRAVASSSLFDLSYLKHSLSQ 58
DB 126 LLAALRLAEPEGAPPAEPVQAPPAVPAAPPAEPAPPAQASASSDITDDEFEQLDLALQ 185
OY 59 QSEPLRLVLVVTLRRIQASMAPAALPPVSPSPAASVDNLLASDAALSASMASL 118
DB 186 GDE-----APASVAAPPAAPAGDEISD-----APEAL 214
OY 119 LEDSHITGLSQAPQPLADEGP-PGRSTGGAAPSLGALDLGPRATGCLLDD--GLEGLFE 175
DB 215 LDQL-HGKGFVPPAVSAEPAQVPAEAVEPAAPAAAGD-DISDDEFEALDLDELHG-KGKFG 271
OY 176 DIDSMYDNEMLWAPASEGLKPGPE-DGPKKEAPELDEAEIDYLDMDVLYG 224
DB 272 DVPEAGTPAAPAAAAAPAAAPAEQKAPAAAGGEISDDEFESLDELHG 321

RESULT 4
US-08-795-868-14
; Sequence 14, Application US/08795868
; Patent No. 5846773
; GENERAL INFORMATION:
; APPLICANT: Lee, Mu-En
; APPLICANT: Hsieh, Chung-Ming
; TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
; TITLE OF INVENTION: AND STRIPED-SPECIFIC MUSCLE CELL ISOFORMS AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,868
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/494,577
; FILING DATE: 22-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/032001
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: 617-542-5070
 TELEFAX: 617-542-8906
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 661 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-08-795-868-14

Query Match 8.6%; Score 104; DB 2; Length 661;
 Best Local Similarity 21.9%; Pred. No. 0.18;

Matches 73; Conservative 28; Mismatches 104; Indels 128; Gaps 15;

QY 7 KKKREEEKEPLAVSW-----WIDPGHA-AVAQAPPAVAS-SSL 45
 DB 109 RRRSLERSDPPAPLPWVPLRKARSLDQPKSERGAPMGCTPGASQELNAPGSVAERRRL 168
 QY 46 FDLSTVLKHLHSLOQSEP-----DLR-----HLYLVVNTLR----- 75
 DB 169 FQOKAASLDERTRQNSPASPDLERFAQELGRTRRSTRRELYVSHESLRTATIQARAPSPRE 228
 QY 76 -----RQASMAPAALPPVSPPAAPSVADNLIASSDA----- 110
 DB 229 PGEPLFSRPTPKTSRAVSPAAPSPSSAEKPGDEPGRPRSRGPA RTEPGEPPQE 288
 QY 111 -----LSASMASLLEDLSHIEGLSQAPQPLAD-----EGPPGRSIGGAAPSGALD 156
 DB 289 VRRDOFPILTRSAIQ-----ECRSVPVPPAADPPEARTKAPGRK---REPPAQA VR 338
 QY 157 LGGPATGCLLDGLEGLEFDIDTSMYDNELMAPASGLKPG-EDGP----- 202
 DB 339 FLPMWT-----PGLBS--AAVPTLEKNRAGPEAEKRLRGRGPEDEGPMGPMDRGARSG 391
 QY 203 ----GKEAPELDEALDYIMDVLTGQALERP 231
 DB 392 KGRRAPTSPELESSDSY---VSAGEEPLERP 421

RESULT 5

US-09-303-069-14
 ; Sequence 14, Application US/09303069A
 ; Patent No. 6350592
 ; GENERAL INFORMATION:
 ; APPLICANT: Hsieh, Chung-Ming
 ; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
 ; TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
 ; FILE REFERENCE: 05433/039001
 ; CURRENT APPLICATION NUMBER: US/09/303,069A
 ; CURRENT FILING DATE: 1999-04-30
 ; EARLIER APPLICATION NUMBER: US 09/134,250
 ; EARLIER FILING DATE: 1998-08-14
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 14
 ; LENGTH: 661
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-303-069-14

Query Match 8.6%; Score 104; DB 4; Length 661;
 Best Local Similarity 21.9%; Pred. No. 0.18;

Matches 73; Conservative 28; Mismatches 104; Indels 128; Gaps 15;

QY 7 KKKREEEKEPLAVSW-----WIDPGHA-AVAQAPPAVAS-SSL 45
 DB 109 RRRSLERSDPPAPLPWVPLRKARSLDQPKSERGAPMGCTPGASQELNAPGSVAERRRL 168
 QY 46 FDLSTVLKHLHSLOQSEP-----DLR-----HLYLVVNTLR----- 75

DB 169 FQOKAASLDERTRQNSPASPDLERFAQELGRTRRSTRRELYVSHESLRTATIQARAPSPRE 228
 QY 76 -----RQASMAPAALPPVSPPAAPSVADNLIASSDA----- 110
 DB 229 PGEPLFSRPTPKTSRAVSPAAPSPSSAEKPGDEPGRPRSRGPA RTEPGEPPQE 288
 QY 111 -----LSASMASLLEDLSHIEGLSQAPQPLAD-----EGPPGRSIGGAAPSGALD 156
 DB 289 VRRDOFPILTRSAIQ-----ECRSVPVPPAADPPEARTKAPGRK---REPPAQA VR 338
 QY 157 LGGPATGCLLDGLEGLEFDIDTSMYDNELMAPASGLKPG-EDGP----- 202
 DB 339 FLPMWT-----PGLBS--AAVPTLEKNRAGPEAEKRLRGRGPEDEGPMGPMDRGARSG 391
 QY 203 ----GKEAPELDEALDYIMDVLTGQALERP 231
 DB 392 KGRRAPTSPELESSDSY---VSAGEEPLERP 421

RESULT 6

US-09-134-250-14
 ; Sequence 14, Application US/09134250B
 ; Patent No. 639753
 ; GENERAL INFORMATION:
 ; APPLICANT: Hsieh, Chung-Ming
 ; TITLE OF INVENTION: SINGLE GENE ENCODING AORTIC-SPECIFIC AND STRIATED-SPECIFIC
 ; TITLE OF INVENTION: MUSCLE CELL ISOFORMS AND USES THEREOF
 ; FILE REFERENCE: 05433/038001
 ; CURRENT APPLICATION NUMBER: US/09/134,250B
 ; CURRENT FILING DATE: 1998-08-14
 ; EARLIER APPLICATION NUMBER: 1997-02-06
 ; EARLIER FILING DATE: 1997-02-06
 ; EARLIER APPLICATION NUMBER: US 08/494,577
 ; EARLIER FILING DATE: 1995-06-22
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 14
 ; LENGTH: 661
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-134-250-14

Query Match 8.6%; Score 104; DB 4; Length 661;
 Best Local Similarity 21.9%; Pred. No. 0.18;

Matches 73; Conservative 28; Mismatches 104; Indels 128; Gaps 15;

QY 7 KKKREEEKEPLAVSW-----WIDPGHA-AVAQAPPAVAS-SSL 45
 DB 109 RRRSLERSDPPAPLPWVPLRKARSLDQPKSERGAPMGCTPGASQELNAPGSVAERRRL 168
 QY 46 FDLSTVLKHLHSLOQSEP-----DLR-----HLYLVVNTLR----- 75
 DB 169 FQOKAASLDERTRQNSPASPDLERFAQELGRTRRSTRRELYVSHESLRTATIQARAPSPRE 228
 QY 76 -----RQASMAPAALPPVSPPAAPSVADNLIASSDA----- 110
 DB 229 PGEPLFSRPTPKTSRAVSPAAPSPSSAEKPGDEPGRPRSRGPA RTEPGEPPQE 288
 QY 111 -----LSASMASLLEDLSHIEGLSQAPQPLAD-----EGPPGRSIGGAAPSGALD 156
 DB 289 VRRDOFPILTRSAIQ-----ECRSVPVPPAADPPEARTKAPGRK---REPPAQA VR 338
 QY 157 LGGPATGCLLDGLEGLEFDIDTSMYDNELMAPASGLKPG-EDGP----- 202
 DB 339 FLPMWT-----PGLBS--AAVPTLEKNRAGPEAEKRLRGRGPEDEGPMGPMDRGARSG 391
 QY 203 ----GKEAPELDEALDYIMDVLTGQALERP 231
 DB 392 KGRRAPTSPELESSDSY---VSAGEEPLERP 421

RESULT 7


```
RESULT 10
US-09-568-480-5
; Sequence 5, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-5
```

```
Query Match      8.6%; Score 103.5; DB 4; Length 7257;
Best Local Similarity 24.2%; Pred. No. 5.1;
Matches 66; Conservative 24; Mismatches 98; Indels 85; Gaps 10;

QY 31 AAVAQAPPAVASSSLEFDL--SVYKLHHSIQOSEPDRLHL--VLVYNTL----- 74
DB 3365 AAKANYAHLEAASGLASLTKAVLALRHEQIPAPPELGEINPMLPWMTLVVAVPRKAVPWG 3424
QY 75 -----RR-----IQASMAPAALPPVSPAPASVADNLLASSDAALSASMAS 117
DB 3425 RGAAPRAVSAFGLSGTIVHYVLEAPVEEPAPAPARVELYVLSAKSAALDAAAR 3484
QY 118 LLEDUS-HIE-----GLSQAPPL-----ADGGPPGRSIG 146
DB 3485 LSAHLSAHPLELSIGDVAFSLATTRSPMEHRLAATTSREALRGALDAAAOQKTPGAVRG 3544
QY 147 GAAPSLGALDLGPATGCLLDGLEGLEF-----DIDTSMYDNEIMAPASEGLKPG 197
DB 3545 KAVSSRGKLAFLFTGGGAQMPGKRGLEYETWPAFRAPRCVALFPREIDQPLREVMAA- 3603
QY 198 PEDGPKKEAPELDEA-----ELDYLMDEV 222
DB 3604 ---APGLAQAAARLDQTAAYAPALFALFALVALAAL 3633
```

```
RESULT 11
US-09-568-486-5
; Sequence 5, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
```

```
; ORGANISM: Sorangium cellulosum
US-09-568-486-5
Query Match      8.6%; Score 103.5; DB 4; Length 7257;
Best Local Similarity 24.2%; Pred. No. 5.1;
Matches 66; Conservative 24; Mismatches 98; Indels 85; Gaps 10;

QY 31 AAVAQAPPAVASSSLEFDL--SVYKLHHSIQOSEPDRLHL--VLVYNTL----- 74
DB 3365 AAKANYAHLEAASGLASLTKAVLALRHEQIPAPPELGEINPMLPWMTLVVAVPRKAVPWG 3424
QY 75 -----RR-----IQASMAPAALPPVSPAPASVADNLLASSDAALSASMAS 117
DB 3425 RGAAPRAVSAFGLSGTIVHYVLEAPVEEPAPAPARVELYVLSAKSAALDAAAR 3484
QY 118 LLEDUS-HIE-----GLSQAPPL-----ADGGPPGRSIG 146
DB 3485 LSAHLSAHPLELSIGDVAFSLATTRSPMEHRLAATTSREALRGALDAAAOQKTPGAVRG 3544
QY 147 GAAPSLGALDLGPATGCLLDGLEGLEF-----DIDTSMYDNEIMAPASEGLKPG 197
DB 3545 KAVSSRGKLAFLFTGGGAQMPGKRGLEYETWPAFRAPRCVALFPREIDQPLREVMAA- 3603
QY 198 PEDGPKKEAPELDEA-----ELDYLMDEV 222
DB 3604 ---APGLAQAAARLDQTAAYAPALFALFALVALAAL 3633
```

```
RESULT 12
US-09-568-472-5
; Sequence 5, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-5
```

```
Query Match      8.6%; Score 103.5; DB 4; Length 7257;
Best Local Similarity 24.2%; Pred. No. 5.1;
Matches 66; Conservative 24; Mismatches 98; Indels 85; Gaps 10;

QY 31 AAVAQAPPAVASSSLEFDL--SVYKLHHSIQOSEPDRLHL--VLVYNTL----- 74
DB 3365 AAKANYAHLEAASGLASLTKAVLALRHEQIPAPPELGEINPMLPWMTLVVAVPRKAVPWG 3424
QY 75 -----RR-----IQASMAPAALPPVSPAPASVADNLLASSDAALSASMAS 117
DB 3425 RGAAPRAVSAFGLSGTIVHYVLEAPVEEPAPAPARVELYVLSAKSAALDAAAR 3484
QY 118 LLEDUS-HIE-----GLSQAPPL-----ADGGPPGRSIG 146
DB 3485 LSAHLSAHPLELSIGDVAFSLATTRSPMEHRLAATTSREALRGALDAAAOQKTPGAVRG 3544
QY 147 GAAPSLGALDLGPATGCLLDGLEGLEF-----DIDTSMYDNEIMAPASEGLKPG 197
DB 3545 KAVSSRGKLAFLFTGGGAQMPGKRGLEYETWPAFRAPRCVALFPREIDQPLREVMAA- 3603
```

OY 198 PEDGPKREAPDELDEA-----ELDYIMDYL 222
DB 3604 ---APGLAQAARLDQDTAYAPALPALEYALAL 3633

RESULT 13

US-09-567-899-5
Sequence 5, Application US/09567899

Patent No. 6383787

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Liqon, James

APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross

APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

FILE REFERENCE: 4-30582A

CURRENT APPLICATION NUMBER: US/09/567,899

PRIOR APPLICATION NUMBER: 09/335,409

PRIOR FILING DATE: 1999-06-17

NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 5

LENGTH: 7257

TYPE: PRT

ORGANISM: Sorangium cellulosum

US-09-567-899-5

Query Match 8.6%; Score 103.5; DB 4; Length 7257;

Best Local Similarity 24.2%; Pred. No. 5.1;

Matches 66; Conservative 24; Mismatches 98; Indels 85; Gaps 10;

OY 31 AAVAQAPPAVASSSLFDL--SVLKLNHSLOOSEPDRLRH--VLVYNTL----- 74

DB 3365 AAKAVAHLEAASGLASLKAIVLALRHQIPRAQPELGLNPHLPNTLPVAVPRKAYPMG 3424

OY 75 -----R-----IQASNAAPAAALPPVPSPPAPSVADNLLASDASASMAS 117

DB 3425 RGARRRRGVSAGFGLSGTNVHVLEAEVEBPAPAPRYELVYLSSAKSAALDAAAR 3484

OY 118 LLEDLS-HIE-----GLSOAPOP-----ADEGPPGRSIG 146

DB 3485 LSAHLSAIPBELSLGVAASLATTSPMEHRLAIAITTSREALGALDAAAOQKTPGCAVRG 3544

OY 147 GAAPSLGALDLIGPATGCLLDGLGLEFE-----DIDTSMYDNELMAPASEGLKPG 197

DB 3545 KAVSSRGKLAFLFTGGAGMPGMRGLYETWPAFREARDRCVALFDRIDQPLREVMA- 3603

OY 198 PEDGPKREAPDELDEA-----ELDYIMDYL 222

DB 3604 ---APGLAQAARLDQDTAYAPALPALEYALAL 3633

RESULT 14

US-08-857-076-57

Sequence 57, Application US/08857076C

Patent No. 6225120

GENERAL INFORMATION:

APPLICANT: Ruvkun, Gary

APPLICANT: Kimura, Koutarou

APPLICANT: Patterson, Garth

APPLICANT: Ogg, Scott

APPLICANT: Paradis, Suzanne

APPLICANT: Tissenbaum, Heidi

APPLICANT: Morris, Jason

APPLICANT: Kowsek, Allison

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR

FILE REFERENCE: 00786/351001

CURRENT APPLICATION NUMBER: US/08/857,076C

CURRENT FILING DATE: 1997-05-15

NUMBER OF SEQ ID NOS: 114
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 57
LENGTH: 655
TYPE: PRT
ORGANISM: Homo sapiens
US-08-857-076-57

Query Match 8.3%; Score 100; DB 3; Length 655;

Best Local Similarity 36.5%; Pred. No. 0.42;

Matches 35; Conservative 9; Mismatches 44; Indels 8; Gaps 3;

OY 74 LRRIQASNAAPAAALPPVPSPPAA--PSYADNLLASDAAASLLEDSHIEGSLQA 131

DB 27 LPRPEFSQNSATSSPAPSGAANPDAAAGLPSAASAAVSADPMNSLLEESDEPQA 86

OY 132 PQLADEGPPGRSIGAPSLGAL--DLIGPATGCL 165

DB 87 PGVA-----AAVMAAATAATGCGDQGFEGACL 118

RESULT 15

US-09-345-473E-40

Sequence 40, Application US/09345473E

Patent No. 6558903

GENERAL INFORMATION:

APPLICANT: Hodge, Martin

TITLE OF INVENTION: NO. 6558903el Kinases and Uses Thereof

FILE REFERENCE: 35800/183781

CURRENT APPLICATION NUMBER: US/09/345,473E

CURRENT FILING DATE: 1999-06-30

NUMBER OF SEQ ID NOS: 62

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 40

LENGTH: 1601

TYPE: PRT

ORGANISM: C. elegans

US-09-345-473E-40

Query Match 8.3%; Score 100; DB 4; Length 1601;

Best Local Similarity 25.1%; Pred. No. 1.4;

Matches 64; Conservative 29; Mismatches 92; Indels 70; Gaps 14;

OY 33 VAQAPPAVASSSLFDLSTL-KLNHSLOOSEPDRLHVLVYNTLRIQASNAAPAAALP-PV 90

DB 1017 VPSVPPVVSNGTL-NLEAPKQTPSATNQNDTQSSSTASTATVSETPATVHTPISV 1075

OY 91 PSPPAAPSVAD-----NLASSDAL-----SASMASLLEDLSHIEGSLQAPO--PLAD 137

DB 1076 PAPVOERPLVIDHSDVLTQLDSELKRVSGVSHSASPVSYSLSL-----SMTPTIPLAC 1130

OY 138 EGPPGRSIGAAPSLGALDLIG-----PATGCLLD--DGLGLFEDIDT----- 179

DB 1131 QTVPA-SIGQAPAVIAAHAAASLIPNASVSPSPRLDAETGLAGLHEKLEALKMEQDRE 1189

OY 180 SMYDNELMAPASEG-----LK-----PEDEGPKREAPDELDELDYIM 219

DB 1190 DMGDAIGTTTTDKDEIPDITLKGALGKVIHADGRETTPMPPHDPLTDA----- 1243

OY 220 DVLVGTQALERRPPG 234

DB 1244 -----STQQLISPSNP 1254

Search completed: August 19, 2003, 09:37:02
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:35:04 ; Search time 57 Seconds
(without alignments)
542.399 Million cell updates/sec

Title: US-09-701-675A-3
Perfect score: 1210
Sequence: 1 MLSKGLKRRKEEKEEPLA.....YLMVDLVGTQALRRPPGR 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 492763 seqs, 131003257 residues
Total number of hits satisfying chosen parameters: 492763

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCOT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCOT_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCOT_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1206	99.7	236	10	US-09-731-872-412
2	168	13.9	254	9	US-09-925-302-596
3	168	13.9	254	15	US-10-106-698-4456
4	165	13.6	241	14	US-10-076-069-4
5	163.5	13.5	157	14	US-10-106-698-5516
6	152	12.6	237	14	US-10-076-069-2
7	115	9.5	255	10	US-09-764-846-139
8	115	9.5	255	15	US-10-091-483-139
9	113	9.3	983	9	US-09-912-917-2
10	112	9.3	713	15	US-10-157-031-415
11	110	9.1	1259	15	US-10-260-715-8
12	108.5	9.0	803	15	US-10-156-761-12958
13	107	8.8	1317	14	US-09-963-896-7
14	104	8.6	113	14	US-10-076-069-6
15	104	8.6	459	9	US-09-789-561-97

16	104	8.6	661	14	US-10-160-865-14	Sequence 14, Appl
17	103.5	8.6	1567	9	US-09-835-232-2	Sequence 2, Appl1
18	103.5	8.6	7257	14	US-10-014-717-5	Sequence 5, Appl1
19	102.5	8.5	674	15	US-10-156-761-9618	Sequence 9618, Ap
20	102	8.4	1516	15	US-10-060-036-166	Sequence 166, App
21	101	8.3	433	10	US-09-801-368-38	Sequence 38, Appl
22	101	8.3	1497	15	US-10-157-031-128	Sequence 128, App
23	100	8.3	655	9	US-09-205-658-57	Sequence 57, Appl
24	100	8.3	655	9	US-09-844-353A-57	Sequence 57, Appl
25	100	8.3	1601	9	US-09-862-027-40	Sequence 40, Appl
26	99.5	8.2	117	14	US-10-076-069-7	Sequence 7, Appl1
27	99.5	8.2	559	15	US-10-156-761-11383	Sequence 11383, A
28	99.5	8.2	682	15	US-10-156-761-12088	Sequence 12088, A
29	99.5	8.2	1232	12	US-10-017-161-2396	Sequence 2396, Ap
30	99.5	8.2	4471	15	US-10-205-032-31	Sequence 10, Appl
31	98.5	8.1	300	15	US-10-156-761-8940	Sequence 8940, Ap
32	97.5	8.1	830	15	US-10-211-962-32	Sequence 32, Appl
33	97	8.0	903	15	US-10-156-761-11093	Sequence 11093, A
34	96.5	8.0	1244	15	US-10-156-761-7652	Sequence 7652, Ap
35	96	7.9	641	15	US-10-156-761-7944	Sequence 7944, Ap
36	96	7.9	1298	10	US-09-825-288A-2	Sequence 2, Appl
37	95	7.9	982	9	US-09-841-132-176	Sequence 176, App
38	95	7.9	1006	9	US-09-862-027-80	Sequence 190, App
39	95	7.9	1265	10	US-10-211-962-31	Sequence 80, Appl
40	94.5	7.8	830	15	US-10-078-770-194	Sequence 31, Appl
41	94	7.8	457	15	US-10-156-761-11646	Sequence 11646, A
42	93.5	7.7	582	15	US-10-027-923-5	Sequence 5, Appl1
43	93.5	7.7	2971	15	US-10-146-473-5	Sequence 50, Appl1
44	93.5	7.7	2971	15	US-10-156-761-11403	Sequence 11403, A
45	93	7.7	267	15		

ALIGNMENTS

US-09-731-872-412
Sequence 412, Application US/09731872
Patent No. US20020102604A1
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, Jean Baptiste
APPLICANT: Bouguenierec, Lydie
TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
FILE REFERENCE: 78 US3 REG
CURRENT APPLICATION NUMBER: US/09/731, 872
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169, 629
PRIOR FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: US 60/187, 470
PRIOR FILING DATE: 2000-03-06
NUMBER OF SEQ ID NOS: 482
SOFTWARE: Patent.pm
SEQ ID NO 412
LENGTH: 236
TYPE: PRT
ORGANISM: Homo sapiens
US-09-731-872-412

Query Match 99.7%; Score 1206; DB 10; Length 236;
Best Local Similarity 99.6%; Pred. No. 2.4e-93;
Matches 235; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLSKGLKRRKEEKEEPLAVSWMDPCHAAVVAQAPVAVASSLFDLSVTLKHSLOOS 60
|||||
Db 1 MLSKGLKRRKEEKEEKEPLAVSWMDPCHVVAQAPVAVASSLFDLSVTLKHSLOOS 60
|||||

QY 61 EEDLRHLVYVNTLRRIQASMAAALPVPSPPAAPVAVDNLASSDAALSASMASTLE 120
|||||
Db 61 EEDLRHLVYVNTLRRIQASMAAALPVPSPPAAPVAVDNLASSDAALSASMASTLE 120
|||||

QY 121 DLSHREGLSQAQPLADEPGRSTGGAPASGALDLCATGCLLDGCLDFEDIDTS 180
|||||

Db 121 DLSHIGSLQAQPPPLADDEGPPGRSLTGGAAPSLGALDILGPAATGCLLDGEGLEFIEDITS 180

QY 181 MYNDELMA PASSEGLKRGPEDEGRGKEAPBLDEAELDYLMDYLVGTQALERPPGGR 236

Db 181 MYNDELMA PASSEGLKRGPEDEGRGKEAPBLDEAELDYLMDYLVGTQALERPPGGR 236

RESULT 2
US-09-925-302-596
; Sequence 596, Application US/09925302
; Patent No. US20020044941A1
JOURNAL INFORMATION

Query Match 13.9%; Score 168; DB 9; Length 254;
Best Local Similarity 27.7%; Pred. No. 2.3e-06;
Matches 73; Conservative 39; Mismatches 90; Indels 62; Gaps 13;

```

RESULT 3
US-10-106-698-4456
; Sequence 4456, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: col. and colon cancer associated polynucleotides and polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280

```

```

: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ. ID NOS: 8564
: SOFTWARE: Patentltn Ver. 3.0
: SEQ. ID NO 4456
: LENGTH: 254
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: MISC_FEATURE
: LOCATION: (105)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: US-10-106-698-4456

```

Query Match	13.9%	Score 168;	DB 15;	Length 254;
Best Local Similarity	27.7%	Pred. No. 2.3e-06;		
Matches 73; Conservative	39;	Mismatches 90;	Indels 62;	Gaps 13

```

1 RESULT 4
2 US-10-076-069-4
3 : Sequence 4, Application US/10076069
4 : Publication No. US2002017721A1
5 : GENERAL INFORMATION:
6 : APPLICANT: JURECIC, ROLAND
7 : TITLE OF INVENTION: HEPV, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEV
8 : FILE REFERENCE: 39532-1176599
9 : CURRENT APPLICATION NUMBER: US/10/076,069
10 : CURRENT FILING DATE: 2002-02-15
11 : PRIOR APPLICATION NUMBER: US 60/268,923
12 : PRIOR FILING DATE: 2001-02-16
13 : NUMBER OF SEQ ID NOS: 11
14 : SOFTWARE: PatentIn version 3.0
15 : SEQ ID NO 4
16 : LENGTH: 241
17 : TYPE: PRT
18 : ORGANISM: Homo sapiens
19 : US-10-076-069-4

```

[illegible]

Db 106 WGOEGHAPAPGLCD-GHTGPGVSDLCPTVSAQAPRHLOSSAWEMDGPRENKSGFHKSLDQ 164
QY 154 ---ALDLGPATCLDLDLGLFEDIDISMYD-NELMAPASEGLKPGEDG----- 201
Db 165 IFETLETKNPS--C-----MEELFSDVDSPIYDLDVVLGMGAGARPGCEGLAPAT 217
QY 202 PGKEAPELDEALDYLMQVYGT 225
Db 218 PGSSSCSKSDGLGELDHVEILVET 241

RESULT 5
US-10-106-698-5516
; Sequence 5516, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106, 698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5516
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (55)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (132)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5516

Query Match 13.5%; Score 163.5; DB 15; Length 157;
Best Local Similarity 32.6%; Pred. No. 3e-06;
Matches 58; Conservative 27; Mismatches 48; Indels 45; Gaps 10;

QY 62 PDLRHLVYVNTLRIOAS--MAPAAPLPVSPAPASVADNLLASSDAISASMASLL 119
Db 11 PSIRRVVLIHNTLQQLQALRLAPALPPEPL-----FLGEEDPSLSAXIGSTIL 60
QY 120 EDL-SHIEGLSQAPQ---LADGPPGRSIGGAAPSIGALDLDGPATGCLDGLGL 173
Db 61 RELDTSMDSTEPQNVPTPLGONEVPP-----QPDVFLEAL--SSRYIGDSGLDGF 111
QY 174 FEDITSMDNLMAPASGLKPGED-----GCKEKAPELDEALDYLMQVYGT 225
Db 112 FLDIDTSAVEKE--PA---RAPEPXMHLFCAPGSWE---WNELDHIMEIILGS 157

RESULT 6
US-10-076-069-2
; Sequence 2, Application US/10076069
; Publication No. US20020177214A1
; GENERAL INFORMATION:
; APPLICANT: JURECIC, ROLAND
; APPLICANT: NACHTMAN, RONALD
; TITLE OF INVENTION: HEP, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVELO
; FILE REFERENCE: 39532-176399
; CURRENT APPLICATION NUMBER: US/10/076, 069
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/268,923
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-076-069-2

Query Match 12.6%; Score 152; DB 14; Length 237;
Best Local Similarity 26.8%; Pred. No. 4.7e-05;
Matches 70; Conservative 35; Mismatches 96; Indels 60; Gaps 10;

QY 1 MSLGKLRKEEBEKEPLAVDSWLDPGHAAVAQAPPAVASSLFLDLSVLKLIHSLQGS 60
Db 1 MFARGLKRRYGDQEE---GVGFGFVPSYSLQRQ-----SLIDMSLVKQLCHMLV 48
QY 61 EPDLRHLVYVNTLRIOASMAP---AALPPVSPAPASVADNLLASS--DAALS 114
Db 49 EPMLCSTVLIANTVRIOEMSGODGVHGMAPQNDRAVERLVSTELLCTRVGAEEER 108
QY 115 MASLLEDLSHIEGLSQAPQPLADEGPPGRSIGAAP-----SLGALDLDGPATGCL-- 165
Db 109 PAPELEDAPLQNSVSLP-----IYGSAPGCRNPQSSSLMEMDSPENKSGPQK 156
QY 166 -LD-----DGLGLFEDIDITSMD-NELMAPASEGLKPGEDG-----PGK 204
Db 157 SLDQIFETLENKSSVEELFSDVDSYYDLDVVLGMGSGKRSICNGLEGPAATPPP 216
QY 205 EEAPELDEALDYLMQVYGT 225
Db 217 SSTCKSDLAELDHVEILVET 237

RESULT 7
US-09-764-846-139
; Sequence 139, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 139
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (19)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-846-139

Query Match 9.5%; Score 115; DB 10; Length 255;
Best Local Similarity 26.6%; Pred. No. 0.065;
Matches 68; Conservative 27; Mismatches 83; Indels 79; Gaps 14;

QY 28 PGHAAVAQAP-----AVASSLFLDLSVLKLIHSLQOSEPDLEHLVYVNT 73
Db 21 PGGAFAASLPDGDATRGGGGLYSAKAPNEMWRCY-----RRKLSLWFLPLPINA 72
QY 74 LRR--IQASMAPAALPPVSPAPASVADNLLASSDAA-ISASMASLLEDLSHIEGLSQ 130

```

Db      73 VRGDFGAGQRC-----PLPGAPAAASALVDHLQVDDAAYPLGLPRHLNLAVDAAH----- 122
Qy      131 APOPLADEGPPRSIGGAAP-----SLGALD-LIGPATGCL---LDDGLEGLF 174
Db      123 -----RLREGRNGRAGSGAAPRLRVRAQGGTAGHTVAADVLELPVHGCAQVQGRGTG-- 176
Qy      175 EDIDTSMYDNELMAPASEGLKRP--ED-----GPGKEAPEL--DEAEIDYLM 220
Db      177 -----ARLDREMHITDAEGIGLGPRLQEESEFGSGPLGPIQLTLPLLFEGCNBPVQD 231
Qy      221 VLVG--TQALERPPEP 234
Db      232 VLLGLPSAVLDRPPLP 247

RESULT 8
US-10-091-483-139
; Sequence 139, Application US/10091483
; Publication No. US20030049650A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PFI212C1
; CURRENT APPLICATION NUMBER: US/10/091,483
; NUMBER OF SEQ ID NOS: 348
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 139
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (18)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (19)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (39)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-10-091-483-139

Query Match
Best Local Similarity 9.5%; Score 115; DB 15; Length 255;
Matches 68; Conservative 27; Mismatches 83; Indels 78; Gaps 14;

Qy      28 PGHAUAQAAP-----AVASSSLFDLSVLKLNHSLOQSEPDRLHVLVNT 73
Db      21 PGPGAASALPGDGTGCGGLYSAKAPNFMWRCV-----RRRLSLNPLFLPINA 72
Qy      74 LRR--IQASMAAALPVPSPRAAPSVADULASSDA--LSASNASLLEDLSHTEGLSQ 130
Db      73 VPRGDFGAGQRC-----PLPGAPAAASALVDHLQVDDAAYPLGLPRHLNLAVDAAH----- 122
Qy      131 APOPLADEGPPRSIGGAAP-----SLGALD-LIGPATGCL---LDDGLEGLF 174
Db      123 -----RLREGRNGRAGSGAAPRLRVRAQGGTAGHTVAADVLELPVHGCAQVQGRGTG-- 176
Qy      175 EDIDTSMYDNELMAPASEGLKRP--ED-----GPGKEAPEL--DEAEIDYLM 220
Db      177 -----ARLDREMHITDAEGIGLGPRLQEESEFGSGPLGPIQLTLPLLFEGCNBPVQD 231
Qy      221 VLVG--TQALERPPEP 234
Db      232 VLLGLPSAVLDRPPLP 247

```

```

RESULT 9
US-09-912-917-2
; Sequence 2, Application US/09912917
; Patent No. US20020028487A1
; GENERAL INFORMATION:
; APPLICANT: Lathague, Nicholas B
; APPLICANT: Shikama, No. US20020028487A1ko
; TITLE OF INVENTION: The University Court of the University of Glasgow
; FILE REFERENCE: AHB/95731500
; CURRENT APPLICATION NUMBER: US/09/912,917
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 09/311,236
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: GB 9818235.5
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-912-917-2

Query Match
Best Local Similarity 9.3%; Score 113; DB 9; Length 983;
Matches 61; Conservative 26; Mismatches 84; Indels 72; Gaps 11;

Qy      3 SKGLKRKEEEREPPLAVDSWMDPGHAUAQAAP-----PAAVSSSLFDLSVLKLNHS 56
Db      94 SPGPRKSOAMAGGSPRARSRLKDPGPRGAPGPESEPLRSPARAKAS----- 141
Qy      57 LQSEPDRLHVLVNTLRIOASWAPAAALPVPSPRAAPSVADNLLASDAALSMA 116
Db      142 -----PLRRAESNDALASATPPRPAPVPVSVRVASGAVS----- 181
Qy      117 SLEDLSHIEGL--SQADQPLAD--EGPPRSIGGAAPSLGALDLGPATGCLDDGLEGL 173
Db      182 --EEIEVLEWREDEADQPLPDSQPP-----SAAELSSAEEC-----SWAGL 223
Qy      174 FEDIDTSMYDNELMAPASEGLKRP-----GPDGPKR-----DEAPRLDAEID--YLM 221
Db      224 FSPDRLRAVHQQLCSVNSQ--LEPCLPVPPEPSSGMMTVLFGAGAPMTBQETDALCYOLOV 282
Qy      222 LVG 224
Db      283 YLG 285

RESULT 10
US-10-157-031-415
; Sequence 415, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequence
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 415
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-415

Query Match
Best Local Similarity 9.3%; Score 112; DB 15; Length 715;

```



```

; REFERENCE/DOCKET NUMBER: PF-0527 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1317 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 2459993
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-963-896-7

Query Match      8.6%; Score 104; DB 10; Length 1317;
Best Local Similarity 23.7%; Pred. No. 2.4;
Matches 58; Conservative 20; Mismatches 79; Indels 88; Gaps 10;

OY 78 QASMAPAALPPVSP-----PAAPSVADN-----LIASSDA 109
DB 746 EGSAEPQLPSPVSPSCGASLPSEASAPDILPAPTPAAGSVTVPEPAPTESSGS 805
OY 110 ALSASMSLDELDSIEGLSQAPQPLADEGPPGKSGIC-----GAAPSGALDL 157
DB 806 SLGGEAPG-SEDEDTTEATSGVFTDLSGDPHTKSGIVPALRSLOKQVGPDSLDSIDI 864
OY 158 LGPAT--GC-----LLDDGL-----EGLFEDIDTSMYDNELMAPAS 191
DB 865 PSSASDGCCEVLSPASAGPPGQPRAVDSGYDTENYESPEVYLKAHSSSEPAFGEFAS 924
OY 192 EGLKPGPBD-----GPKKEAPELDEA--ELDYLDVIVG-----TQALER 230
DB 925 EGSPGPDPLSVSLGSLSKSPYRDSAYFSDLDLAESEPTGPKHSGIQDSQKEODIRS 984
OY 231 PPGPG 235
DB 985 PPSPG 989

RESULT 14
US-10-076-069-6
; Sequence 6, Application US/10076069
; Publication No. US20020177214A1
; GENERAL INFORMATION:
; APPLICANT: JURCEIC, ROLAND
; APPLICANT: NACHTMAN, RONALD
; TITLE OF INVENTION: HEP, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVELOPMENT
; FILE REFERENCE: 39532-176599
; CURRENT APPLICATION NUMBER: US/10/076,069
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/268,923
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-076-069-6

Query Match      8.6%; Score 104; DB 14; Length 113;
Best Local Similarity 34.1%; Pred. No. 0.19;
Matches 28; Conservative 16; Mismatches 26; Indels 12; Gaps 2;

OY 1 MLSGLKRRKEEKEEPLAVDSWMLDPSGHAAVAQAPPAVASSSLPFLSVLKLHSLQGS 60
DB 1 MEANGLRKRGDDE-----GVGEGYVPSISLQRO-----SLDMSLVKLDLCHMLV 48
OY 61 EPDLRHLVVLVNTLRRIQASMA 82
DB 49 EPNLCRSVLTANTVRAIOIEKMS 70
```

```

RESULT 15
US-09-789-561-97
; Sequence 97, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2001-02-22
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 97
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (345)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-789-561-97

Query Match      8.6%; Score 104; DB 9; Length 459;
Best Local Similarity 24.6%; Pred. No. 1.1;
Matches 56; Conservative 18; Mismatches 68; Indels 86; Gaps 11;

OY 24 WMLDPGHAANAQAPPAVASSSLPFLSVLKLHSLQSEPDRLHLVL----- 69
DB 93 WRCCPGHSGVSCCEVAASASLSEPM-----WSGSTRRRALPTAFSGCLNCSKY 142
OY 70 --VNTLRRIQASMAPAAL--PPVSPPAAPSVADNLLASDALSMSMSLDELDSHIE 126
DB 143 SELTERLKVLEAKKMTLTVIEQVPPTPATP-----EDPAPLV 180
OY 127 GLSQAPQPLADEGPPGKSGISGAAPSLGALDLGPGATGCLDDGLEGLFEDIDTSMYDNEL 186
DB 181 G-----PPPA--QGSPPG--GGLDQVYGMGLPGP--TGPKGDAAGSG----- 217
OY 187 WAPASEGLKPGEPDEGPKKEAPELDEAELDYLDVIVGTQALERPPGP 234
DB 218 --PMGMRGPPGPQGPQPG---SPGRAGA-----VGTPOERGPPPGP 251

Search completed: August 19, 2003, 09:44:59
Job time : 58 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 15:06:01 ; Search time 310 Seconds
(without alignments)
8644.552 Million cell updates/sec

Title:	US-09-701-675A-8
Perfect score:	1192
Sequence:	1 gtgagccgagtcagactg.....yggaatctcaaaaaaaaa 1192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

```
Searched:      1517243 seqs, 1124081882 residues
Total number of hits satisfying chosen parameters: 3034486
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database : Published_Applications_NA:*

- 1: /cgn2_6/prodata/2/pubnpa/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pubnpa/PC1_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/2/pubnpa/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pubnpa/US07_NEW_PUB.seq.*
- 5: /cgn2_6/prodata/2/pubnpa/PC1US_PUBCOMB.seq.*
- 6: /cgn2_6/prodata/2/pubnpa/US08_NEW_PUB.seq.*
- 7: /cgn2_6/prodata/2/pubnpa/US08_PUBCOMB.seq.*
- 8: /cgn2_6/prodata/2/pubnpa/US09_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pubnpa/US09_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pubnpa/US09_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pubnpa/US09C_NEW_PUB.seq.*
- 12: /cgn2_6/prodata/2/pubnpa/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/2/pubnpa/US10_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/2/pubnpa/US10_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pubnpa/US10_NEW_PUB.seq.*
- 16: /cgn2_6/prodata/2/pubnpa/US60_NEW_PUB.seq.*
- 17: /cgn2_6/prodata/2/pubnpa/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Description			
No.	Score	Query Length	ID	
1	1188.8	99.7	1219	US-09-731-872-171
2	1188.8	99.7	1219	US-09-876-997-171
3	258	21.6	258	US-09-796-692-5565
4	258	21.6	258	US-10-040-862-5565
5	120.8	10.1	165	US-09-728-455-520
6	46.2	3.9	1227	US-10-156-761-2641
7	46.2	3.9	9025608	US-10-156-761-1
8	44.6	3.7	2082	US-10-076-069-3
9	44.6	3.7	2629	US-10-037-670-773
10	43	3.6	396	US-09-825-294-167
11	43	3.6	396	US-09-970-966-167
12	43	3.6	396	US-10-212-677-167
13	43	3.6	2038	US-09-925-302-153
14	43	3.6	2038	US-10-106-698-179
15	43	3.6	9025608	US-10-156-761-1
16	42.2	3.5	1049	US-10-140-672-358

C	17	42.2	3.5	1049	12	US-10-141-761-358	Sequence 358, App
C	18	42.2	3.5	1049	14	US-10-123-155-358	Sequence 358, App
C	19	42.2	3.5	1049	15	US-10-146-731-358	Sequence 358, App
C	20	42.2	3.5	1428	14	US-10-145-413-15	Sequence 15, Appl
C	21	42.2	3.5	2082	13	US-10-076-069-1	Sequence 1, Appl
C	22	42	3.5	10056	14	US-10-156-761-411	Sequence 411, App
C	23	42	3.5	100000	14	US-10-156-761-15103	Sequence 15103, A
C	24	41.8	3.5	1413	14	US-10-145-415-5	Sequence 5, Appl
C	25	41.8	3.5	1683	14	US-10-156-761-5824	Sequence 5824, App
C	26	41.4	3.5	2949	14	US-10-187-267A-12	Sequence 12, Appl
C	27	40.8	3.5	36521	14	US-10-187-267B-1	Sequence 1, Appl
C	28	41.4	3.4	1387	11	US-09-756-481C-1	Sequence 1, Appl
C	29	40.6	3.4	378	14	US-10-156-761-332	Sequence 332, App
C	30	40.6	3.4	474	11	US-09-918-995-3967	Sequence 3967, App
C	31	40.4	3.4	2940	14	US-10-156-761-3172	Sequence 3172, App
C	32	40.2	3.4	268	11	US-09-373-658-58	Sequence 58, Appl
C	33	39.4	3.3	1344	14	US-10-156-761-1108	Sequence 1108, App
C	34	39	3.3	1836	14	US-10-156-761-2659	Sequence 2659, App
C	35	39	3.3	13862	11	US-09-764-891-5477	Sequence 5477, App
C	36	39	3.3	13862	11	US-09-764-891-10204	Sequence 10204, A
C	37	39	3.3	13862	14	US-10-255-428-1003	Sequence 1003, App
C	38	38.8	3.3	735	13	US-10-027-632-158772	Sequence 158772, App
C	39	38.8	3.3	31871	14	US-09-764-847-1403	Sequence 1403, App
C	40	38.8	3.3	31871	14	US-10-092-154-1403	Sequence 1403, App
C	41	38.4	3.2	2109	10	US-09-738-973-153	Sequence 153, App
C	42	38.4	3.2	2109	10	US-09-854-133-153	Sequence 153, App
C	43	38.4	3.2	2109	14	US-10-144-649A-153	Sequence 153, App
C	44	38.4	3.2	3377	13	US-10-027-632-113850	Sequence 113850, App
C	45	38.2	3.2	2968	9	US-09-815-915-1	Sequence 1, Appl

ALIGNMENTS

```

RESULT 1
; US-09-731-872-171
; Sequence 171, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bouqueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNs ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78.US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: patent.pm
; SEQ ID NO 171
; LENGTH: 1219
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..867
; US-09-731-872-171

Query Match          99.7%; Score 1188.8; DB 10; Length 1219;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY	Db	QY
1	GTGAGCGCGGATGAGACTGCGCTCTCGCGACCCAGCGCGGGTTTCCGAGGACAGCCAA	60
21	GTGAGCGCGGAGTACAGTACTGCGCTCTCGCGACCCAGCGCGGGTTTCCGAGGAGAGGCCAA	80
61	CAAGCATCTCTGCGCGCGCGCTTCTCTGATTTGTTGAGGTGAGCGTACCTCTTCGTTTCGA	120
81	CAAGCATCTCTGCGCGCGCGCTTCTCTGATTTGTTGAGGTGAGCGTACCTCTTCGTTTCGA	140

121 TTGGCCGCTAGTACCAAGATGCTGAGCAAGGGCTCTGAAGCGGAACGGAGAGAGAGA 180
141 TTGGCCGCTAGTACCAAGATGCTGAGCAAGGGCTCTGAAGCGGAACGGAGAGAGAGA 200
181 GGAGAGAGAACCTCTGGGAGATGAGTCTGCTGGGCTGATATCTGGCCACGACGGGTGGC 240
201 GGAGAGAGAACCTCTGGGAGATGAGTCTGCTGGGCTGATATCTGGCCACGACGGGTGGC 260
241 ACAGGACACCCCGGCGCTGAGCTTACCTCTTTTACCTCTGAGAGCTCAAGCTCA 300
261 ACAGGACACCCCGGCGCTGAGCTTACCTCTTTTACCTCTGAGAGCTCAAGCTCA 320
301 CCACAGCTGAGAGAGATGAGCGGACCTGGCGACCTGGTGTCTGAGACACCT 360
321 CCACAGCTGAGAGAGATGAGCGGACCTGGCGACCTGGTGTCTGAGACACCT 380
361 GGGGGGCAATCCAGGGCTCATGGGCAAGCGGGCTCCCTGGCACCTGGTGTCTGAGCCAC 420
381 GGGGGGCAATCCAGGGCTCATGGGCAAGCGGGCTCCCTGGCACCTGGTGTCTGAGCCAC 440
421 TGACACCCCGGAGTGTGGCTGACAACTTACTGGCAAGCTGAGACGCTTTCAGCCCTC 480
441 TGACACCCCGGAGTGTGGCTGACAACTTACTGGCAAGCTGAGACGCTTTCAGCCCTC 500
481 CATGGCCAGCCTCTGAGAGACCTGACCAATTGAGGGCTCTGAGTCAAGCTCCCAAC 540
501 CATGGCCAGCCTCTGAGAGACCTGACCAATTGAGGGCTCTGAGTCAAGCTCCCAAC 560
541 CTGGCAGACGAGGGGGCCACGAGCGCTGAGCATGGGGGAGAGAGCGCCACGCTGGGTGC 600
561 CTGGCAGACGAGGGGGCCACGAGCGCTGAGCATGGGGGAGAGAGCGCCACGCTGGGTGC 620
601 CTGGACCTGTGTGGGCCCGGACCTGCTGTCTACTGAGCAGATGGGCTTGAAGGCTGTT 660
621 CTGGACCTGTGTGGGCCCGGACCTGCTGTCTACTGAGCAGATGGGCTTGAAGGCTGTT 680
661 TGAGGATTTGACACCTCTATGTATGACAAATGAACTTTGGGACACGCTCTGAGGGCT 720
681 TGAGGATTTGACACCTCTATGTATGACAAATGAACTTTGGGACACGCTCTGAGGGCT 740
721 CAACACGAGCGCTGAGATGGGCGGCAAGAGAGAACTCCGAGCTGAGAGAGAGAGAGA 780
741 CAACACGAGCGCTGAGATGGGCGGCAAGAGAGAACTCCGAGCTGAGAGAGAGAGAGA 800
781 ATTGACTACTCATGATGATGTGTGTGGGCAACAGGCACTGAGAGAGAGAGAGAGAG 840
801 ATTGACTACTCATGATGATGTGTGTGGGCAACAGGCACTGAGAGAGAGAGAGAGAG 860
841 AGGGGCTGAGAGCGCTGAGATGGGCGGCAAGAGAGAACTCCGAGCTGAGAGAGAGAG 900
861 AGGGGCTGAGAGCGCTGAGATGGGCGGCAAGAGAGAACTCCGAGCTGAGAGAGAGAG 920
901 GAGCAACTGTCTCTGAG 960
921 GAGCAACTGTCTCTGAG 980
961 CACTTTGAG 1020
981 CACTTTGAG 1040
1021 CTGGCAGGAGGAG 1080
1041 CTGGCAGGAGGAG 1100
1081 GAATTCCTGAG 1140
1101 TAATTCCTGAG 1160
1141 CCAATTCAGATTTTCAATTAAGCCAGTCTGGGAAATCTCAAAAAA 1192
1161 CCAATTCAGATTTTCAATTAAGCCAGTCTGGGAAATCTCAAAAAA 1212

RESULT 2
US-09-876-997-171
; Sequence 171, Application US/09876997
; Publication No. US20030152921A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bouqueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNA ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78 USA, CIP
; CURRENT APPLICATION NUMBER: US/09/876, 997
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 09/731, 872
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/187, 470
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: US 60/169, 629
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 171
; LENGTH: 1219
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..867
US-09-876-997-171
Query Match 99.7%; Score 1188.8; DB 12; Length 1219;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 GTGAGCCGAGTCAAGAACTGCTCGGAGCCAGCGCGGGTTCCGAGAGAGACCA 60
21 GTGAGCCGAGTCAAGAACTGCTCGGAGCCAGCGCGGGTTCCGAGAGAGACCA 80
61 CAAGGATGCTGCGCGCGCGCTTCCTGATTTGTTGGTGGCTTACTCTTCCTTGA 120
81 CAAGGATGCTGCGCGCGCGCTTCCTGATTTGTTGGTGGCTTACTCTTCCTTGA 140
121 TTGGCCGCTAGTACCAAGATGCTGAGCAAGGGCTCTGAAGCGGAAGAGAGAGA 180
141 TTGGCCGCTAGTACCAAGATGCTGAGCAAGGGCTCTGAAGCGGAAGAGAGAGA 200
181 GGAGAGAGAACCTCTGGGAGATGAGTCTGCTGGGCTGATATCTGGCCACGACGGGTGGC 240
201 GGAGAGAGAACCTCTGGGAGATGAGTCTGCTGGGCTGATATCTGGCCACGACGGGTGGC 260
241 ACAGGACACCCCGGCGCTGAGCTTACCTCTTTTACCTCTGAGAGCTCAAGCTCA 300
261 ACAGGACACCCCGGCGCTGAGCTTACCTCTTTTACCTCTGAGAGCTCAAGCTCA 320
301 CCACAGCTGAGAGAGATGAGCGGACCTGGCGACCTGGTGTCTGAGACACCT 360
321 CCACAGCTGAGAGAGATGAGCGGACCTGGCGACCTGGTGTCTGAGACACCT 380
361 GGGGGGCAATCCAGGGCTCATGGGCAAGCGGGCTCCCTGGCACCTGGTGTCTGAGCCAC 420
381 GGGGGGCAATCCAGGGCTCATGGGCAAGCGGGCTCCCTGGCACCTGGTGTCTGAGCCAC 440
421 TGACACCCCGGAGTGTGGCTGACAACTTACTGGCAAGCTGAGACGCTTTCAGCCCTC 480
441 TGACACCCCGGAGTGTGGCTGACAACTTACTGGCAAGCTGAGACGCTTTCAGCCCTC 500
481 CATGGCCAGCCTCTGAGAGACCTGACCAATTGAGGGCTCTGAGTCAAGCTCCCAAC 540
501 CATGGCCAGCCTCTGAGAGACCTGACCAATTGAGGGCTCTGAGTCAAGCTCCCAAC 560
541 CTGGCAGACGAGGGGGCCACGAGCGCTGAGCATGGGGGAGAGAGCGCCACGCTGGGTGC 600
561 CTGGCAGACGAGGGGGCCACGAGCGCTGAGCATGGGGGAGAGAGCGCCACGCTGGGTGC 620

OY	601	CTTGAGACCTGGTGGGCCACGACACATGGGCTGCTACTGGACATGGGCTGGAGGGCCGTGT	660
Db	621	CTTGAGACCTGGTGGGCCACGACACATGGGCTGCTACTGGACATGGGCTGGAGGGCCGTGT	680
OY	661	TGAGGATATTTGACACCTCTATGTATGACAAATGAACCTTGGGCACACAGCTCTGGAGGCGCT	720
Db	681	TGAGGATATTTGACACCTCTATGTATGACAAATGAACCTTGGGCACACAGCTCTGGAGGCGCT	740
OY	721	CAAAACGAGGCTTATGATGGGCGGGGACAGGAGAAAGCTCCGAGAGCTGGACGAGGCGGA	780
Db	741	CAAAACGAGGCTTATGATGGGCGGGGACAGGAGAAAGCTCCGAGAGCTGGACGAGGCGGA	800
OY	781	ATTGAGACACTCAATNGATGTGCGTGGTGGGACACACAGGACATGGAGGAGCCGCCAGGGGCC	840
Db	801	ATTGAGACACTCAATNGATGTGCGTGGTGGGACACACAGGACATGGAGGAGCCGCCAGGGGCC	860
OY	841	AGGGCGCTGAGCCCTCGTGTGATGATGGTGTCTGTATCTGAACTAGAGCTGCTGGCGTG	900
Db	861	AGGGCGCTGAGCCCTCGTGTGATGATGGTGTCTGTATCTGAACTAGAGCTGCTGGCGTG	920
OY	901	GACCACTGTCCTGAAAAGACACAGCTGGCTCCCTAGTACAGAGAAACAGGGCTTGGGC	960
Db	921	GACCACTGTCCTGAAAAGACACAGCTGGCTCCCTAGTACAGAGAAACAGGGCTTGGGC	980
OY	961	CACCTTGGAGAGAGAAATCTAGTCCTGCGGGCAATTCAACCCGCTCCCTCGTCTCAGGG	1020
Db	981	CACCTTGGAGAGAGAAATCTAGTCCTGCGGGCAATTCAACCCGCTCCCTCGTCTCAGGG	1040
OY	1021	CTGGCAGGGGAGAGCTGGAATTACCCCTAGTAGTGAATAGACAGGCTCTGGTGGGAGCT	1080
Db	1041	CTGGCAGGGGAGAGCTGGAATTACCCCTAGTAGTGAATAGACAGGCTCTGGTGGGAGCT	1100
OY	1081	GAATTCCTCGGCCCTGGGGCTATAGCTTGGGCTTCCTCTCTGATACGGGAAGAGACC	1140
Db	1101	TAATTCCTCGGCCCTGGGGCTATAGCTTGGGCTTCCTCTCTGATACGGGAAGAGACC	1160
OY	1141	CCAATCAGATTTTCAAATTAACCCAGCTCTGGGAAATCTCAAAAAAAAAA	1192
Db	1161	CCAATCAGATTTTCAAATTAACCCAGCTCTGGGAAATCTCAAAAAAAAAA	1212

RESULT 3
US-09-796-692-5565/c
: Sequence 5565, Application US/09796692
: Publication No. US20020198362A1
: GENERAL INFORMATION:
: APPLICANT: Gaiger, Alexander
: APPLICANT: Algate, Paul A.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
: TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
: FILE REFERENCE: 2077_001200
: CURRENT APPLICATION NUMBER: US/09/796,692
: CURRENT FILING DATE: 2001-03-01
: PRIOR APPLICATION NUMBER: 60/186,126
: PRIOR FILING DATE: 2000-03-01
: PRIOR APPLICATION NUMBER: 60/190,479
: PRIOR FILING DATE: 2000-03-17
: PRIOR APPLICATION NUMBER: 60/200,545
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: 60/200,303
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,779
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,999
: PRIOR FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 60/202,084
: PRIOR FILING DATE: 2000-05-04
: PRIOR APPLICATION NUMBER: 60/206,201
: PRIOR FILING DATE: 2000-05-22
: PRIOR APPLICATION NUMBER: 60/218,950
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: 60/222,903

```

: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: 60/223,416
: PRIOR FILING DATE: 2000-08-04
: PRIOR APPLICATION NUMBER: 60/223,378
: PRIOR FILING DATE: 2000-08-07
: NUMBER OF SEQ ID NOS: 9597
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5565
: LENGTH: 258
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-796-692-5565

```

	Query Match	21.6%	Score 258	DB 10	Length 258
	Best Local Similarity	100.0%	Pred. No. 5e-66		
	Matches	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0
QY	927	CTGGCTTCCCTAGTACAGAGAACAAGGCTTTGGGGCACCCTTTGGAGACAGAAATCTAGTCC	986		
DB	258	CTGGCTTCCCTAGTACAGAGAACAAGGCTTTGGGGCACCCTTTGGAGAGACGAATCTAGTCC	199		
QY	987	TGGGCAACTTCACATCCGTCCTCCTGTCTCAGGGCTGGCAGGGGAGCCTGGAATTACC	1046		
DB	198	TGGGCAACTTCACATCCGTCCTCCTGTCTCAGGGCTGGCAGGGGAGCCTGGAATTACC	139		
QY	1047	CCTAGTATGGAATGACAGGGCTGTGGTGGGAGCTGAATTCCTCGCCCTGGGGGTATAGC	1106		
DB	138	CCTAGTATGGAATGACAGGGGTGTGGTGGGAGCTGAATTCCTCGCCCTGGGGGTATAGC	79		
QY	1107	TTGGGCGCTTCCTCTCTGATACGGGGAAGAGACCCCATTCAGATTTTCAATTTAAACC	1166		
DB	78	TTGGGCGCTTCCTCTCTGATACGGGGAAGAGACCCCATTCAGATTTTCAATTTAAACC	19		
QY	1167	AGTCTCGGAATCTCAA	1184		
DB	18	AGTCTCGGAATCTCAA	1		

```

1      RESULT 4
2      US-10-040-862-5565/c
3      ; Sequence 5565, Application US/10040862
4      ; Publication No. US20030078396A1
5      ;
6      ; GENERAL INFORMATION:
7      ;
8      ; APPLICANT: Gaiger, Alexander
9      ; APPLICANT: Algate, Paul A.
10     ; APPLICANT: Mannion, Jane
11     ; APPLICANT: Reltter, Marc
12     ;
13     ; APPLICANT: Corixa Corporation
14     ;
15     ; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
16     ; TITLE OF INVENTION: Hematological Malignancies
17     ;
18     ; FILE REFERENCE: 014058-013520US
19     ;
20     ; CURRENT APPLICATION NUMBER: US//10/040,862
21     ;
22     ; CURRENT FILING DATE: 2001-11-06
23     ;
24     ; PRIOR APPLICATION NUMBER: US 60/186,126
25     ;
26     ; PRIOR FILING DATE: 2000-03-01
27     ;
28     ; PRIOR APPLICATION NUMBER: US 60/190,479
29     ;
30     ; PRIOR FILING DATE: 2000-03-17
31     ;
32     ; PRIOR APPLICATION NUMBER: US 60/200,545
33     ;
34     ; PRIOR FILING DATE: 2000-04-27
35     ;
36     ; PRIOR APPLICATION NUMBER: US 60/200,303
37     ;
38     ; PRIOR FILING DATE: 2000-04-28
39     ;
40     ; PRIOR APPLICATION NUMBER: US 60/200,779
41     ;
42     ; PRIOR FILING DATE: 2000-04-28
43     ;
44     ; PRIOR APPLICATION NUMBER: US 60/200,999
45     ;
46     ; PRIOR FILING DATE: 2000-05-01
47     ;
48     ; PRIOR APPLICATION NUMBER: US 60/202,084
49     ;
50     ; PRIOR FILING DATE: 2000-05-04
51     ;
52     ; PRIOR APPLICATION NUMBER: US 60/206,201
53     ;
54     ; PRIOR FILING DATE: 2000-05-22
55     ;
56     ; PRIOR APPLICATION NUMBER: US 60/218,950
57     ;
58     ; PRIOR FILING DATE: 2000-07-14
59     ;
60     ; PRIOR APPLICATION NUMBER: US 60/222,903
61     ;
62     ; PRIOR FILING DATE: 2000-08-03
63     ;

```



```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARDO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match          3.9%; Score 46.2; DB 14; Length 9025608;
Best Local Similarity 45.3%; Pred. No. 0.028;
Matches 168; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY      483 TGCCAGACCTCTGAGAGACCTCAGCCACATTGAGGGCCCTGAGTCCAGGCTCCCAACCT 542
Db      3253368 TCGTCCGCGCCCTCGCCGAGCCGATCAGTTGACCGGCAACCCAGCTCGCTGCACG 3253427

QY      543 TGGCAGACGAGGGGCCACGACCGCGTACGATTCGGGGAGAGAGGCCAGCTGGTGCCT 602
Db      3253428 GCGATGACGATGCGCCCTCTGGGCGGTGCGCTCCAGCTTGCTGAGACGATGCGGTGATA 3253487

QY      603 TGGACCTGCTGGGCCACGACACCTGCTCTACTGACGATGGGCTTGAAGGCTCTTTG 662
Db      3253488 TTGACGACCTCGCGCGAAGACACCTGCTGACACGCGCTTTCGCCCGGTGGGCTCG 3253547

QY      663 AGGATATTGACACCTTATGATGATGACATGAACTTTGGGACACGACCTCTGAGGGCTCA 722
Db      3253548 AGGACGACGAGTACCTCTGTCGACGCGGGGCTGCTTTCGACGACGCGCTTGCCCC 3253607

QY      723 AACCCAGCCCTGAGGATGGGCCGCGCAAGAGAGAGCTCCGAGCTGGAGAGCGCGAAT 782
Db      3253608 AGCTGCTCCATGAGCCCGGTGTTGTGACGCGGCCCGCGGTGTCGATGAGAGAGACG 3253667

QY      783 TGGACTACCTCATGATGATGCTGCTGGGACACACAGGACACTGAGAGACCGCGGGGCGAG 842
Db      3253668 TCGGCGCCCTCCCTCGCATGCGCTCTTGACGGGCTCGAAGCGCATTCAGCGCGGTGCG 3253727

QY      843 GCGCGTGAAGCC 853
Db      3253728 CCCTCGGGTCC 3253738

RESULT 8
US-10-076-069-3
; Sequence 3, Application US/10076069
; Publication No. US20020177214A1
; GENERAL INFORMATION:
; APPLICANT: JURECIC, ROLAND
; APPLICANT: NACHTMAN, RONALD
; TITLE OF INVENTION: HEPF, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVELOPMENT
; FILE REFERENCE: 39532-176599
; CURRENT APPLICATION NUMBER: US/10/076,069
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/268,923
; PRIOR FILING DATE: 2001-02-16
```

```

; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 2082
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75)..(797)
US-10-076-069-3

Query Match          3.7%; Score 44.6; DB 13; Length 2082;
Best Local Similarity 50.7%; Pred. No. 0.0072;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY      269 TCCCTTTGACCTCAGTGTCAAGCTCCACACAGCCTGAGAGAGTGGCCGAG 328
Db      180 TCGCTCTGAGCATGTCTCTGTGTGAAGTTGCACAGTTCGACATGCTTGTGGAGCCCAAC 239

QY      329 CTGCGCACCTGTGTGCTGTGTGAACACTCTGCGGCGCATCCAGGCGTCCATGACACCC 388
Db      240 CTGTGCGCGTCACTCTCATTTGCCAACAGCGTCCGGAGATCCAAAGAGAGATGACGACG 299

QY      389 GCGGCTGCCCTGCGCACCTGTGCTTAGCCACCTGACAGCCCCAGTGTGCTGACAACTTA 448
Db      300 GATGGGACGTGGCGCACAGTGGCACCCACAGCTGACAGAGCGGCGCGCTGCACCGCTTG 359

QY      449 CTGCAAGCTCGGAGCGTGCCTTTGACGCT 479
Db      360 GTCTCCACGAGATCTCTGTGCTGCGTGCAGGCT 390

RESULT 9
US-10-037-270-773
; Sequence 773, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhou, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aiding J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Dimauc, Radoje T.
; TITLE OF INVENTION: NO. US20030104529A1el Nucleic Acids and
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/10/037,270
; PRIOR FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: PL_FL-genes Version 1.0
; SEQ ID NO 773
; LENGTH: 2629
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73)..(798)
US-10-037-270-773
```

Query Match 3.7%; Score 44.6; DB 14; Length 2629;
Best Local Similarity 50.7%; Pred. No. 0.0077;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

269 TCCCTCTTTGACCTCTCAGTGTCTCAAGCTCCACACAGCTTGACAGAGTGAAGCCGGAC 328
178 TCGCTCCGAGACATGTCTGTGTGAAGTTGACGCTTTGGACATGTGTGTGAAGCCAAAC 237
QY 329 CTGGCGGACCTGTGTGTGTGTGTGAACACTCTGCGCGGACATCCAGCCCTTCATGCGACCC 388
DB 238 CTGTCCCGCTCAGTCTCATTTGTCACACAGGTCGCGAGATCCCAAGAGAGATGACGAG 297
QY 389 GCGGCTGCGCTGCACCTGT 448
DB 298 GATGGAGAGTGGCGGACAGTGGCACACCCAGGCTTGACAGAGGCGGCGCGCTGCAGCGCTTG 357
QY 449 CTGGCAAGCTGCGAGCGTGTGCTTTTCAGCCT 479
DB 358 GTCTCCACGAGATCTGTGCGGTGACGCT 388

RESULT 10
US-09-825-294-167
; Sequence 167, Application US/09825294
; Patent No. US20020004491A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Algate, Paul A.
; APPLICANT: Filing, Steven P.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.484C5
; CURRENT APPLICATION NUMBER: US/09/825,294
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 167
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-825-294-167

Query Match 3.6%; Score 43; DB 9; Length 396;
Best Local Similarity 50.2%; Pred. No. 0.013;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

269 TCCCTCTTTGACCTCTCAGTGTCTCAAGCTCCACACAGCTTGACAGAGTGAAGCCGGAC 328
DB 173 TCGCTCCGAGACATGTCTGTGTGAAGTTGACGCTTTGGACATGTGTGTGAAGCCAAAT 232
QY 329 CTGGCGGACCTGTGTGTGTGTGTGAACACTCTGCGCGGACATCCAGCCCTTCATGCGACCC 388
DB 233 CTGTCCCGCTCAGTCTCATTTGTCACACAGGTCGCGAGATCCCAAGAGAGATGACGAG 292
QY 389 GCGGCTGCGCTGCACCTGT 448
DB 293 GATGGAGAGTGGCGGACAGTGGCACACCCAGGCTTGACAGAGGCGGCGCGCTGCAGCGCTTG 352
QY 449 CTGGCAAGCTGCGAGCGTGTGCTTTTCAGCCT 479
DB 353 GTCTCCACGAGATCTGTGCGGTGACGCT 383

RESULT 11
US-09-970-966-167
; Sequence 167, Application US/09970966
; Patent No. US20020173638A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Molesh, David Alan
; APPLICANT: Filing, Steven P.

APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.484C6
; CURRENT APPLICATION NUMBER: US/09/970,966
; CURRENT FILING DATE: 2001-10-02
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-970-966-167

Query Match 3.6%; Score 43; DB 10; Length 396;
Best Local Similarity 50.2%; Pred. No. 0.013;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

269 TCCCTCTTTGACCTCTCAGTGTCTCAAGCTCCACACAGCTTGACAGAGTGAAGCCGGAC 328
DB 173 TCGCTCCGAGACATGTCTGTGTGAAGTTGACGCTTTGGACATGTGTGTGAAGCCAAAT 232
QY 329 CTGGCGGACCTGTGTGTGTGTGTGAACACTCTGCGCGGACATCCAGCCCTTCATGCGACCC 388
DB 233 CTGTCCCGCTCAGTCTCATTTGTCACACAGGTCGCGAGATCCCAAGAGAGATGACGAG 292
QY 389 GCGGCTGCGCTGCACCTGT 448
DB 293 GATGGAGAGTGGCGGACAGTGGCACACCCAGGCTTGACAGAGGCGGCGCGCTGCAGCGCTTG 352
QY 449 CTGGCAAGCTGCGAGCGTGTGCTTTTCAGCCT 479
DB 353 GTCTCCACGAGATCTGTGCGGTGACGCT 383

RESULT 12
US-10-212-677-167
; Sequence 167, Application US/10212677
; Publication No. US20030129192A1
; GENERAL INFORMATION:
; APPLICANT: Chenuault, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Fanger, Gary R.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.484C7
; CURRENT APPLICATION NUMBER: US/10/212,677
; CURRENT FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 288
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-212-677-167

Query Match 3.6%; Score 43; DB 15; Length 396;
Best Local Similarity 50.2%; Pred. No. 0.013;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

269 TCCCTCTTTGACCTCTCAGTGTCTCAAGCTCCACACAGCTTGACAGAGTGAAGCCGGAC 328
DB 173 TCGCTCCGAGACATGTCTGTGTGAAGTTGACGCTTTGGACATGTGTGTGAAGCCAAAT 232
QY 329 CTGGCGGACCTGTGTGTGTGTGTGAACACTCTGCGCGGACATCCAGCCCTTCATGCGACCC 388
DB 233 CTGTCCCGCTCAGTCTCATTTGTCACACAGGTCGCGAGATCCCAAGAGAGATGACGAG 292
QY 389 GCGGCTGCGCTGCACCTGT 448
DB 293 GATGGAGAGTGGCGGACAGTGGCACACCCAGGCTTGACAGAGGCGGCGCGCTGCAGCGCTTG 352

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 13:39:16 ; Search time 2460 Seconds
(without alignments)
11776.803 Million cell updates/sec

Title: US-09-701-675A-8

Sequence: 1 gtgagccggaagtcagaactg.....gggaatctcaaaaaaaaaa 1192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :
EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_hlv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	856.6	71.9	1023	12	BM923629	BM923629 AGENCOURT
2	798.8	67.0	908	13	BH856774	BH856774 AGENCOURT
3	786.6	66.0	1062	12	BM558458	BM558458 AGENCOURT
4	781.4	65.6	797	13	BK101521	BK101521 BK101521

5	756.4	63.5	766	13	B0539198	B0539198	AGENCYCOURT
6	751.8	63.1	803	10	BG715050	BE715050	B06763273
7	749.2	62.9	868	13	B0433568	B0433568	AGENCYCOURT
8	744.2	62.4	883	13	BQ878247	B0878247	AGENCYCOURT
9	743.4	62.4	765	13	B0537987	B0537987	AGENCYCOURT
10	742.2	62.3	846	10	BG468137	BG468137	AGENCYCOURT
11	740.8	62.1	782	12	BG8299426	BG8299426	602509712
12	737	61.8	747	14	CD364995	CD364905	UI-H-FT2-
13	733.2	61.5	1037	10	BG402912	BE402912	602418715
14	730.8	61.3	764	13	B0538362	B0538362	AGENCYCOURT
15	727	61.0	949	13	B0557431	B0557431	AGENCYCOURT
16	722.4	60.6	923	13	B0957131	B0957131	AGENCYCOURT
17	716.6	60.1	732	14	CA748763	CA748763	UI-H-FT1-
18	708.8	59.5	878	13	B0690528	B0690528	AGENCYCOURT
19	699.2	58.7	782	12	B1753849	B1753849	603027530
20	693.8	58.2	876	10	BE898247	BE898247	601681014
21	693.2	58.2	882	13	B0690477	B0690477	AGENCYCOURT
22	687.4	57.7	927	13	BQ182938	BQ182938	UI-H-ED1-
23	683.6	57.3	952	13	B0957038	B0957038	AGENCYCOURT
24	682.2	57.2	768	10	BG546883	BE546883	602574065
25	667	56.8	698	13	B0633212	B0633212	UI-H-FT1-
26	662.8	55.6	695	12	BM973302	BM973302	UI-CF-EC1
27	656	55.0	701	10	BE613164	BE613164	601452139
28	656	55.0	1189	11	AK0064022	AK0064022	MUS_muscu
29	655.4	55.0	666	14	CA420820	CA420820	UI-H-FT0-
30	654.8	54.9	677	14	CA431558	CA431558	UI-H-FT1-
31	654.4	54.9	1185	11	AK008303	AK008303	MUS_muscu
32	646	54.4	656	14	CH851800	CH851800	UI-CF-FNO
33	646	54.2	662	14	CB250543	CB250543	UI-CF-FNO
34	637.8	53.5	674	9	A1343377	A1343377	q127e08.x
35	628.2	52.7	655	10	BE898853	BE898853	601682146
36	627.6	52.7	674	13	BQ771685	BQ771685	UI-H-ED1-
37	624.4	52.4	627	12	BM843483	BM843483	K-EST0121
38	612.8	51.4	1187	11	AK002324	AK002324	MUS_muscu
39	610.6	51.2	645	14	CD368900	CD368900	UI-H-FT1-
40	610.2	51.2	787	10	BF032665	BF032665	601453004
41	591	49.6	627	12	BM9419292	BM9419292	AGENCYCOURT
42	580.6	48.7	1016	14	BY704271	BY704271	AGENCYCOURT
43	578	48.5	638	10	BE613003	BE613003	601452139
44	574.2	48.2	688	10	BE271936	BE271936	60141534
45	570.8	47.9	1123	12	B1470179	B1470179	602963868

ALIGNMENTS

RESULT	1
BPM923629	
LOCUS	
DEFINITION	BM923629 1023 bp mRNA linear EST_12-MAR-2002 AGENCOURT_6623809 NIH_MGC_116 Homo sapiens cDNA IMAGE:5759328
ACCESSION	BM923629
VERSION	BM923629
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 1023) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished
TITLE	Contact: Robert Strausberg, Ph.D.
JOURNAL	
COMMENT	

FEATURES
source

Location/Qualifiers
1..1023
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5759328"
/lab_host="DH10B"
/clone.lib="NIH_MGC_116"
/note="Organ: pooled colon, kidney, stomach; Vector:
pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dt primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH_MGC Library."

BASE COUNT 192 a 313 c 318 g 196 t 4 others

ORIGIN

Query Match 71.9%; Score 856.6; DB 12; Length 1023;
Best Local Similarity 96.7%; Pred. No. 4.2e-190;
Matches 904; Conservative 0; Mismatches 28; Indels 3; Gaps 3;

19 TCGCTCTGCGACCCAGCGCGGGTTCGCGAGACAGCCAAAGCGATGCTGCCGCG 78
|||||
5 TTCGGATCGCGACCCAGCGCGGGTTCGCGAGACAGCCAAAGCGATGCTGCCGCG 64
79 CCGTTCTGANTGGTTGGTGGTGGTACCTCTTGTCTGATTGGCCGCTAGTAGCAA 138
|||||
65 CCGTTCTGANTGGTTGGTGGTGGTACCTCTTGTCTGATTGGCCGCTAGTAGCAA 124
139 GATGCTGAGCAAGGCTCTGAAGCGGAACGGGAGAGAGAGAGAGAAAGAACTCTGGC 198
125 GATGCTGAGCAAGGCTCTGAAGCGGAACGGGAGAGAGAGAGAGAAAGAACTCTGGC 184
139 AGTCACATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 258
185 AGTCACATCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 244
259 GGCCCTAGCTCCCTCTTGAACCTCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 318
245 GGCCCTAGCTCCCTCTTGAACCTCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 304
319 TGAGCGGACCTGGCGGACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 378
305 TGAGCGGACCTGGCGGACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 364
379 CATGGACACCGCGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 438
365 CATGGACACCGCGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 424
439 TGACAACCTTACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCT 498
425 TGACAACCTTACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCTGGGACCT 484
499 GGACCTAGCCACATTTGAGGCGCTGAGTCAAGCTCCCAACCTTGGCAGAGAGAGGCGC 558
485 GGACCTAGCCACATTTGAGGCGCTGAGTCAAGCTCCCAACCTTGGCAGAGAGAGGCGC 544
559 ACCAGGCGCTAGCATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 618
545 ACCAGGCGCTAGCATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604
619 AGCCACTGGCTGTACTGAG 678
605 AGCCACTGGCTGTACTGAG 664
679 TATGTATGACAAATGAACTTTGGGACACAGCTCTGAGAGGCTCAAAACAGGAGCCCTGA-GG 737
665 TATGTATGACAAATGAACTTTGGGACACAGCTCTGAGAGGCTCAAAACAGGAGG 724

QY 738 ATGGCCCGGAG 797
|||||
DB 725 ATGGCCCGGAG 784
QY 798 ATGTCTGTGTGGGACACAGGACCTGAGACCGCCGCGGAGAGAGAGAGAGAGAGAGAG 856
|||||
DB 785 ATGTCTGTGTGGGACACAGGACCTGAGACCGCCGCGGAGAGAGAGAGAGAGAGAGAG 844
QY 857 GTGCTGAATGGTTGT 915
|||||
DB 845 GTGCTGAATGGTTGT 904
QY 916 AAAAGACACAGCTGGCTTCCTAGTACAGAGACA 950
|||||
DB 905 AAAAGACACAGCTGGCTTCCTAGTACAGAGACA 939

RESULT 2
B0856974 908 bp mRNA linear EST 16-OCT-2002
LOCUS B0856974
DEFINITION AGENCOURT_10481033 NIH_MGC_107 Homo sapiens cDNA clone
IMAGE:5646774 5', mRNA sequence.
ACCESSION B0856974
VERSION B0856974.1 GI:24041964
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 908)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>

Plate: L1CM2886 row: b column: 22
High quality sequence stop: 717.

FEATURES

location/Qualifiers
1..908
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5646774"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone.lib="NIH_MGC_107"
/note="Organ: breast; Vector: pOT7; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dt priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 182 a 259 c 267 g 170 t 30 others

ORIGIN

Query Match 67.0%; Score 798.8; DB 13; Length 908;
Best Local Similarity 98.4%; Pred. No. 1.4e-176;
Matches 817; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 354 AACTCTGGCGGCGATCCAGGCGTGCATGACACCGCGGCGTCCCTGCACTGTGCTTA 413
|||||
DB 1 AACTCTGGCGGCGATCCAGGCGTGCATGACACCGCGGCGTCCCTGCACTGTGCTTA 60

OY		414	GCCACCGTCGAGGCCGCCAGCTGTGGCTGTGCACATTAAGGGCAAGCCTCGSAGCCTCCCTTT	473
Db		61	GGCCACCTCGACGCCCCCACTGTGGCTGTGCACATTAAGGGCAAGCCTCGSAGCCTCCCTTT	120
OY		474	CAGCCTCCATTGGCCAGCCTCCTGTGGAGAACCTCACACCACATTAAGAGGGCCTTAGTCAAGCTC	533
Db		121	CAGCCTCCATTGGCCAGCCTCCTGTGGAGAACCTCACACCACATTAAGAGGGCCTTAGTCAAGCTC	180
OY		534	CCCAAACCTTGGCAGACGAGGGGCCACACAGGCCGTAGCATCGGGGGAGACGCCCCAGCC	593
Db		181	CCCAAACCTTGGCAGACGAGGGGCCACACAGGCCGTAGCATCGGGGGAGACGCCCCAGCC	240
OY		594	TGGGGCCCTTGAGACTGTGTGGGGCCAGCACTAGTGTGTCTGTGACAGATGGGCTTAGAG	653
Db		241	TGGGGCCCTTGAGACTGTGTGGGGCCAGCACTAGTGTGTCTGTGACAGATGGGCTTAGAG	300
OY		654	GCCGTGTTGAGAGATATTGACACCTCTATATGTATGAACAATGAATCTTGGGACACACCTCTG	713
Db		301	GCCGTGTTGAGAGATATTGACACCTCTATATGTATGAACAATGAATCTTGGGACACACCTCTG	360
OY		714	AGGGCCTTCAAAACAGGCCCTTGAGGATGGGCCGGGCCAAGAGGAAGCTTCCGAGCTGGACG	773
Db		361	AGGGCCTTCAAAACAGGCCCTTGAGGATGGGCCGGGCCAAGAGGAAGCTTCCGAGCTGGACG	420
OY		774	AGGGCGAATTGGAATACTTCACTTCAATGTGTGTGTGGGACACACAGGCACTGGAGCGACCGC	833
Db		421	AGGGCGAATTGGAATACTTCACTTCAATGTGTGTGTGGGACACACAGGCACTGGAGCGACCGC	480
OY		834	CGGGGCCAGGGCGCTGAGCCCTCGTGGCTGGAAATGGTTGTCTGGATATCGAATCAAGCCTG	893
Db		481	CGGGGCCAGGGCGCTGAGCCCTCGTGGCTGGAAATGGTTGTCTGGATATCGAATCAAGCCTG	540
OY		894	CTGGCTGAGACACACTGTCTCTGAAAAGAAGACACAGCTGGCTTCCCTAAGTACAGAGAACAAGG	953
Db		541	CTGGCTGAGACACACTGTCTCTGAAAAGAAGACACAGCTGGCTTCCCTAAGTACAGAGAACAAGG	600
OY		954	CTTGGGGCCACTTTTGGAGAGACAGAATATAGTCTCTGGGCACTTACATCCGCTCCTCTGT	1013
Db		601	CTTGGGGCCACTTTTGGAGAGACAGAATATAGTCTCTGGGCACTTACATCCGCTCCTCTGT	660
OY		1014	CTCAGAGCTGGCAGGGGGAGGCGCTGGAATATCCCCCTAGTGTGATGTGATGACAGGGCTCGGT	1073
Db		661	CTCAGAGCTGGCAGGGGGAGGCGCTGGAATATCCCCCTAGTGTGATGTGATGACAGGGCTCGGT	720
OY		1074	GGGGACGTGAATTCCTCTGGCCCTCGGGGGTCATA-GCTTGGGCTGTCTCTCTGTCTGATACGGG	1132
Db		721	GGGGACGTGAATTCCTCTGGCCCTCGGGGGTCATANGCTTGGGCGTGTCTCTCTGTGATACGGG	780
OY		1133	AAGAGACCCCACATCAGATTTTTCAAATTAAGCAAGCAGTCTCTGGAAATCTC	1182
Db		781	AAGAGACCCCACATCAGATTTTTCAAATTAAGCAAGCAGTCTCTGGAAATCTC	830
RESULT_3				
BMS58458				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				

FEATURES			
source			
<p>CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: Mgc clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LRAM12757 row: j column: 11 High quality sequence stop: 616. Location/Qualifiers</p>			
1. .1062 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5741290" /tissue_type="duodenal adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_idb="NH.MGC.88" /note="Organ: small intestine; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NH.MGC Library." 			
BASE COUNT	204 a	325 c	339 g
ORIGIN		190 t	4 others
Query Match	66.0%;	Score 786.6;	DB 12; Length 1062;
Best Local Similarity	87.6%;	Pred. No. 1.1e-173;	
Matches	922;	Conservative	0; Mismatches 11; Indels 20; Gaps 5
QY	49	GAGGACACCAACGATGCTGGCGCCGCTTCTCGATTGGTTGGGTGGCTACC	108
DB	2	GACGACACCAACGATGCTGGCGCCGCTTCTCGATTGGTTGGGTGGCTACC	61
QY	109	TCTTCGTTCTATTGGCGCTAGTAGACAAGATGCTGAGCAAGGCTCTGAACGGAAACG	168
DB	62	TCTTCGTTCTATTGGCGCTAGTAGACAAGATGCTGAGCAAGGCTCTGAACGGAAACG	121
QY	169	GGAGGAGGAGGAGGAAGAACTCTGGCACTGCACTCTGGTGGCTAGATCTGGGCA	228
DB	122	GGAGGAGGAGGAGGAAGAACTCTGGCACTGCACTCTGGTGGCTAGATCTGGGCA	181
QY	229	CGACGCGGTGGACAGGGACCCCGCGGTGGCGCTAGTCCCTTGGACCTCGACT	288
DB	182	CGACGCGGTGGACAGGGACCCCGCGGTGGCGCTAGTCCCTTGGACCTCGACT	241
QY	289	GCTCAAGCTCACCACAGCCTTGACAGAGTAGAGCGGACCTGCGGACCTGGTGGT	348
DB	242	GCTCAAGCTCACCACAGCCTTGACAGAGTAGAGCGGACCTGCGGACCTGGTGGT	301
QY	349	CGTGAACACTCTGGCGGCATCCAGGCGTCCATGGCACCCGCGGTGCCCTGCCACCTGT	408
DB	302	CGTGAACACTCTGGCGGCATCCAGGCGTCCATGGCACCCGCGGTGCCCTGCCACCTGT	361
QY	409	GCCTTAGCCCACTGAGGCCCAAGTGTGGCTACAACTTACCTGGCAAGCTGGAGGCTGC	468
DB	362	GCCTTAGCCCACTGAGGCCCAAGTGTGGCTACAACTTACCTGGCAAGCTGGAGGCTGC	421
QY	469	CCTTTCACGCTTCATGGCGCAGCCTCTGGAGAACCTCAGCCACATTTGAGGGGCTGAGTCA	528
DB	422	CCTTTCACGCTTCATGGCGCAGCCTCTGGAGAACCTCAGCCACATTTGAGGGGCTGAGTCA	481
QY	529	GGCTTCCCAACCCCTTGGGAGAGAGAGGGGCAACAGCGGTACATCTGGGGGAGCAGCGCC	588
DB	482	GGCTTCCCAACCCCTTGGGAGAGAGAGGGGCAACAGCGGTACATCTGGGGGAGCAGCGCC	541
QY	589	CAGCCTGGGTGCTTGGAGCTGCTGGGCGCCACGACTGCTGTCTACCTGAGATGGGCG	601
DB	542	CAGCCTGGGTGCTTGGAGCTGCTGGGCGCCACGACTGCTGTCTACCTGAGATGGGCG	601
QY	648	TTGAGGGGCTTTTGAAGATATTGACACCTCATATGATAGACAATAAATTGGGGACACAG	707
DB	602	TTGAGGGGCTTTTGAAGATATTGACACCTCATATGATAGACAATAAATTGGGGACACAG	661
QY	708	CCTTCGAGGCGCTCAACACGAGGCCCTTGAAGATGGGCGCGGCAAGAGCAAGCTCCGAGAC	767

LOCUS BU539198 766 bp mRNA linear EST 13-SEP-2002
 DEFINITION AGENCOURT_10215071 NIH_MGC_107 Homo sapiens cDNA clone
 IMAGE:6569886 5', mRNA sequence.
 ACCESSION BU539198
 VERSION BU539198.1 GI:22849639
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 766)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHC2757 row: 0 column: 06
 High quality sequence stop: 705.
 Location/Qualifiers
 1..766
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6569886"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_107"
 /note="Organ: breast; Vector: pORF7; Site:1; EcoRI;
 Site:2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."
 BASE COUNT 169 a 209 c 236 g 152 t
 ORIGIN
 Query Match 63.5%; Score 756.4; DB 13; Length 766;
 Best Local Similarity 99.9%; Pred. No. 1.2e-166;
 Matches 757; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 435 TGGCTGACAACTTACTGCGAAGCTCGAGCGTGCCTTCAGCCTCCATGCGCAGCCTCC 494
 |||||||
 Db 1 TGGCTGACAACTTACTGCGAAGCTCGAGCGTGCCTTCAGCCTCCATGCGCAGCCTCC 60
 OY 495 TGGAGACCTCGACCCACATTGAGAGGCGCTGAGTACGCTCCCAACCTTGGCAGAGAG 554
 |||||||
 Db 61 TGGAGACCTCGACCCACATTGAGAGGCGCTGAGTACGCTCCCAACCTTGGCAGAGAG 120
 OY 555 GGCACACAGGCGCTAGCATCGGGGAGAGAGCGCCAGCTGGGTGCTTGGACCTGCTGG 614
 |||||||
 Db 121 GGCACACAGGCGCTAGCATCGGGGAGAGAGCGCCAGCTGGGTGCTTGGACCTGCTGG 180
 OY 615 GCCCAGCCCACTGGCTGTACTGAGAGATGGGCTTGAGGGCTGTGTTGAGGATATTGACA 674
 |||||||
 Db 181 GCCCAGCCCACTGGCTGTACTGAGAGATGGGCTTGAGGGCTGTGTTGAGGATATTGACA 240
 OY 675 CCTCTATGATGACATGAACTTTGGGACACAGCCTCTGAGAGGCTCAAAACCGGCCCTG 734
 |||||||
 Db 241 CCTCTATGATGACATGAACTTTGGGACACAGCCTCTGAGAGGCTCAAAACCGGCCCTG 300
 OY 735 AGGATAGGCGGGGCAAGAGAACTCCGAGCTGAGAGAGGCGGCAATTGGACTACTCTCA 794
 |||||||
 Db 301 AGGATAGGCGGGGCAAGAGAACTCCGAGCTGAGAGAGGCGGCAATTGGACTACTCTCA 360

FEATURES

source
 1..766
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6569886"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_107"
 /note="Organ: breast; Vector: pORF7; Site:1: EcoRI;
 Site:2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."

OY 795 TGGATGTGCTGTGGGACACAGGACCTGGAGCGACCCGCCGGGCCAGGCGCTGAGCC 854
 |||||||
 Db 361 TGGATGTGCTGTGGGACACAGGACCTGGAGCGACCGCGGGGCCAGGCGCTGAGCC 420
 OY 855 TCGTGTGGAATGTTGTCTGTGATCTGAACCTAGAGCTGCTGCTGACCACTGTC 914
 |||||||
 Db 421 TCGTGTGGAATGTTGTCTGTGATCTGAACCTAGAGCTGCTGCTGACCACTGTC 480
 OY 915 GAAAGACACAGCTGCTGCCCTAGTACAGAGAAAGAGGCTTGGGCACTTTGGAGAGAC 974
 |||||||
 Db 481 GAAAGACACAGCTGCTGCCCTAGTACAGAGAAAGAGGCTTGGGCACTTTGGAGAGAC 540
 OY 975 AGAATCTAGTCTGTGGCAACTTCACATCCGCTCTCTCTGCTGAGGCTGCGAGGAGAGC 1034
 |||||||
 Db 541 AGAATCTAGTCTGTGGCAACTTCACATCCGCTCTCTCTGCTGAGGCTGCGAGGAGAGC 600
 OY 1035 CTGGAAATTACCCCTTGTATGTAATGACAGGCTGTGGGACATGATTCCTGAGCC 1094
 |||||||
 Db 601 CTGGAAATTACCCCTTGTATGTAATGACAGGCTGTGGGACATGATTCCTGAGCC 660
 OY 1095 TGGGTCATAGCTTGGGCTGTCTCTCTGATGAGGGAAGAGACCCCAATCAGATTTT 1154
 |||||||
 Db 661 TGGGTCATAGCTTGGGCTGTCTCTCTGATGAGGGAAGAGACCCCAATCAGATTTT 720
 OY 1155 CAAATTAAAGCCAGTCTGTGGAAATCTCAAAAAAAAAA 1192
 |||||||
 Db 721 CAAATTAAAGCCAGTCTGTGGAAATCTCAAAAAAAAAA 758

RESULT 6

BG715050 803 bp mRNA linear EST 08-MAY-2001
 LOCUS BG715050
 DEFINITION 602673273F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4795699 5',
 mRNA sequence.
 BG715050
 BG715050.1 GI:13993981

ACCESSION

VERSION BG715050.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 803)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM10679 row: b column: 20
 High quality sequence stop: 802.
 Location/Qualifiers
 1..803
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4795699"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_96"
 /note="Organ: Brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site:1: BamHI; Site:2: SalI; XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
 size-selected for average insert size 2.3 kb and
 normalized to R0T 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 150 a 249 c 257 g 147 t

ORIGIN

Query Match 63.1%; Score 751.8; DB 10; Length 803;
Best Local Similarity 98.7%; Pred. No. 1.4e-165;
Matches 789; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 9 GAGTCAGAACTGCTGTCGACGACGAGGCGGGTTTCCGAGGACAGCAACAGCAT 68
D 1
D 4 GGGTCAGAACTGCTGTCGACGACGAGGCGGGTTTCCGAGGACAGCAACAGCAT 63
QY 69 GGTGCGCGCGCGGTTTCTGATTTGTTGGGTGGCTTCTTCTTCTGATTTGGCCG 128
D 64 GGTGCGCGCGCGGTTTCTGATTTGTTGGGTGGCTTCTTCTTCTGATTTGGCCG 123
QY 129 TGTGTGAGAAAGATGTCGACGACGAGGCTGTAAGCGGAAAGGAGGAGGAGGAGG 188
D 124 TGTGTGAGAAAGATGTCGACGACGAGGCTGTAAGCGGAAAGGAGGAGGAGGAGG 183
QY 189 AACCTCTGGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 248
D 184 AACCTCTGGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 243
QY 249 CCCCCGCGGCGGCTTCTGATTTGTTGGGTGGCTTCTTCTTCTGATTTGGCCG 308
D 244 CCCCCGCGGCGGCTTCTGATTTGTTGGGTGGCTTCTTCTTCTGATTTGGCCG 303
QY 309 TGCACAGAGTGTGAGCGGACGCTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTG 368
D 304 TGCACAGAGTGTGAGCGGACGCTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTG 363
QY 369 TGCAGGCGTCCATGTCACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 428
D 364 TGCAGGCGTCCATGTCACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 423
QY 429 CCAGTGTGGCTGACAACTTACGAGCACTGCGAGGCTGCTGCTGCTGCTGCTGCTG 488
D 424 CCAGTGTGGCTGACAACTTACGAGCACTGCGAGGCTGCTGCTGCTGCTGCTGCTG 483
QY 489 GCGTCTGTGAGAGCTTCAACATTTGAGGCGCTGAGTCAAGCTTCCCAACCTTGGAG 548
D 484 GCGTCTGTGAGAGCTTCAACATTTGAGGCGCTGAGTCAAGCTTCCCAACCTTGGAG 543
QY 549 ACAGAGGCGCACAGCGGCTGAGTCTGCGGAGAGCGCCCAAGCTTGGCTTGGAGC 608
D 544 ACAGAGGCGCACAGCGGCTGAGTCTGCGGAGAGCGCCCAAGCTTGGCTTGGAGC 603
QY 609 TGCCTGGGCGCACAGCGGCTGAGTCTGAGAGGCTTGAAGGCTTGTGAGGATA 668
D 604 TGCCTGGGCGCACAGCGGCTGAGTCTGAGAGGCTTGAAGGCTTGTGAGGATA 663
QY 669 TTGACACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 727
D 664 TTGACACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 723
QY 728 GAGCCTGAGATGAGGCGGAGGAGG-AGGAAAGCTCCGAGGCTGAGAGGAGGAGGAG 786
D 724 AGCCTTGAGATGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 783
QY 787 CTACTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 805
D 784 CTACTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 801

RESULT 7 868 bp mRNA linear EST 24-MAY-2002
LOCUS BQ433568
DEFINITION AGNCOURT_7761470 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020198
ACCESSION BQ433568
VERSION BQ433568.1 GI:21172644

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 868)
AUTHORS NIH-MGC <http://mhc.nhl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LLM13223 row: 0 column: 15
High quality sequence stop: 614.
Location/Qualifiers
1. 868
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6020198"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV-Sport6; Site: 1; Not; Site: 2; Salt; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.1 kb. Library constructed by Life Technologies."

BASE COUNT 157 a 265 c 284 g 160 t 2 others

ORIGIN

Query Match 62.9%; Score 749.2; DB 13; Length 868;
Best Local Similarity 98.7%; Pred. No. 5.9e-165;
Matches 818; Conservative 0; Mismatches 4; Indels 7; Gaps 6;

QY 35 GCGCGGGTTTCCGAGAGCAGCAACAGGATGTCGCGCGCTTCTGATTTGGT 94
D 1 GCGCGGGTTTCCGAGAGCAGCAACAGGATGTCGCGCGCTTCTGATTTGGT 60
QY 95 TGTGGGTGGCTTACCTTTCTGTTGATTGGCGGTAAGTGTGAGCAAGATCTGAGCAAGG 154
D 61 TGTGGGTGGCTTACCTTTCTGTTGATTGGCGGTAAGTGTGAGCAAGATCTGAGCAAGG 120
QY 155 CTGAAGGCGAAGCGGAG 214
D 121 CTGAAGGCGAAGCGGAG 180
QY 215 CTAGATCTGGCCACAGAGCGGTGGACAGGACCCCGCGCGGTCTAGCTCCCTC 274
D 181 CTAGATCTGGCCACAGAGCGGTGGACAGGACCCCGCGCGGTCTAGCTCCCTC 240
QY 275 TTGACCTTCAAGTCTCAAGCTTCCACACAGGCTGAGAGAGAGAGAGAGAGAGAG 334
D 241 TTGACCTTCAAGTCTCAAGCTTCCACACAGGCTGAGAGAGAGAGAGAGAGAGAG 300
QY 335 CACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 394
D 301 CACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
QY 395 GCCCTGCACCTGTGCTAGCGCCACGTCAGAGCCCGCGAGTGGGTGCACTTACTGGCA 454
D 361 GCCCTGCACCTGTGCTAGCGCCACGTCAGAGCCCGCGAGTGGGTGCACTTACTGGCA 420
QY 455 AGTCGAGAGCGTGGCTTTGAGCGCTTCATGAGCGGAGCGGCTCTGAGAGAGAGAG 514
D 421 AGTCGAGAGCGTGGCTTTGAGCGCTTCATGAGCGGAGCGGCTCTGAGAGAGAGAG 480
QY 515 GAGGCGCTGAGTCAAGGCTCCCAACCTTGGCAGAGAGAGGCGCACCAAGCGCTAGCATC 574

|||||
Db 481 GAGGGCTAGTCAGGCTCCCAACCCCTTGCGACAGAGGGGGCCACAGCCGTACATC 540
QY 575 GGGGGAGCAGCGCCAGCTGGGTGCTTGACCTCTGGGCCAGCCACTGGCTGTCTA 634
Db 541 GGGGGAGCAGCGCCAGCTGGGTGCTTGACCTCTGGGCCAGCCACTGGCTGTCTA 600
QY 635 CTGGAGCA-TGGGCTTGAGGGCCCTG-TTTGAGGATATTTGACACCTCTATGTATGCAATG 692
Db 601 CTGGAGCATTTGGGCTTGAGGGCCCTGTTTGAGGATATTTGACACCTCTATGTATGCAATG 660
QY 693 AACTTTGGGACACGAGCTTGAGGGCCCTTAACACAGG-CCCTGAGATGGGCC-GGGCAA 750
Db 661 AACTTTGGGACACGAGCTTGAGGGCCCTTAACACAGG-CCCTGAGATGGGCC-GGGCAA 720
QY 751 GGAGAGAGCTCCGAGAGCTGGAGAGGCGCAATTTGACTACTCATG--ATGAGTGGTG 808
Db 721 GGAGAGAGCTCCGAGAGCTGGAGAGGCGCAATTTGACTACTCATGAGATGGTGGTG 780
QY 809 GGCACACAGGCACTGGAGCGACCG-CCGGGGCCAGGGCGCTGAGCCCTC 856
Db 781 GGCACACAGGCACTGGAGCGACCGCGGGGCCAGGGCGCTGAGCCCTC 829

RESULT 8
B0878247 883 bp mRNA linear EST 16-AUG-2002
LOCUS AGENCOURT.8074327 NIH_MGC_110 Homo sapiens CDNA clone IMAGE:6085638
DEFINITION 5', mRNA sequence.
ACCESSION B0878247
VERSION B0878247.1 GI:22270255
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 883)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cygabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
Cloning Strategy: Agencourt Bioscience Corporation
Cloning Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L12M2319 row: f column: 07
High quality sequence stop: 553.
Location/Qualifiers
1. 883

FEATURES
source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6085638"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_110"
/note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAGG(G). Library constructed by
ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

BASE COUNT 163 a 277 c 285 g 158 t
ORIGIN
Query Match 62.4%; Score 744.2; DB 13; Length 883;
Best Local Similarity 98.0%; Pred. No. 8.7e-164;

Matches 785; Conservative 0; Mismatches 13; Indels 3; Gaps 3;
QY 43 TTTCGAGAGACAGCCAAAGAGATGTGCCGCCCTTTCTGATTGGTGTGGGG 102
Db 2 TTTCGAGAGACAGCCAAAGAGATGTGCCGCCCTTTCTGATTGGTGTGGGG 61
QY 103 GCTACCTCTTCTGATTGGCCGCTAGTGAACAAGATGTGAGCAAGGGTCTGAAGC 162
Db 62 GCTACCTCTTCTGATTGGCCGCTAGTGAACAAGATGTGAGCAAGGGTCTGAAGC 121
QY 163 GAAACGGAG 222
Db 122 GAAACGGAG 181
QY 223 TGCCACAG 282
Db 182 TGCCACAG 241
QY 283 CTGAGTGTCAAGCTTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 342
Db 242 CTGAGTGTCAAGCTTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
QY 343 GCTGCTGTGAACACTCTGCGGCGCATCCAGAGAGAGAGAGAGAGAGAGAGAG 402
Db 302 GCTGCTGTGAACACTCTGCGGCGCATCCAGAGAGAGAGAGAGAGAGAGAGAG 361
QY 403 ACCGTGCTAG 462
Db 362 ACCGTGCTAG 421
QY 463 CGCTGCTCTTTCAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 522
Db 422 CGCTGCTCTTTCAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
QY 523 GAGTCAAGCTCTTTCAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 582
Db 482 GAGTCAAGCTCTTTCAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541
QY 583 AGCGCCACAGCTGGAG 642
Db 542 AGCGCCACAGCTGGAG 601
QY 643 TGGGCTTGAG 702
Db 602 TGGGCTTGAG 661
QY 703 ACCAGCTTGAG 761
Db 662 ACCAGCTTGAG 721
QY 762 CGGAGCTTGAG 820
Db 722 CGGAGCTTGAG 781
QY 821 CT-GGAG 840
Db 782 CTGGAG 802

RESULT 9
B0537987 765 bp mRNA linear EST 13-SEP-2002
LOCUS AGENCOURT.10186374 NIH_MGC_107 Homo sapiens CDNA clone
DEFINITION IMAGE:6568492 5', mRNA sequence.
ACCESSION B0537987
VERSION B0537987.1 GI:22848428
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 765)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D.

FEATURES	Location/Qualifiers
source	1. .765

BASE COUNT	169 a	208 c	235 g	152 t	1 others
ORIGIN					

REFERENCE	1 (bases 1 to 782)
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D.

BASE COUNT	144 a	243 c	250 g	145 t
ORIGIN				

D	b67	GATTGGTTGTGSGTGGCTACTCCTTCGTTTGATTTGGCGGTAGTAGCAAGAATGTGAC	126
OY	148	CAAAGGTCTGAAGCGGAAAACGGGAGGAGAGAGAGAGAACCTCTGCAGTGCAC TC	207
D	127	CAAAGGTCTGAAGCGGAAAACGGGAGGAGAGAGAGAGAACCTCTGCAGTGCAC TC	186
OY	208	CTGGTGGCTAATATCCTGGCCACAAGCGGTGGCACAGGCCACCCTCCCCTGGCTCATG	267
D	187	CTGGTGGCTAATATCCTGGCCACAAGCGGTGGCACAGGCCACCCTCCCCTGGCTCATG	246
OY	268	CTCCCTCTTAACCTCTCAGTGTCTCAAGCTCCACACAGCCTTGCAGCAGATGTGAGCCGA	327
D	247	CTCCCTCTTAACCTCTCAGTGTCTCAAGCTCCACACAGCCTTGCAGCAGATGTGAGCCGA	306
OY	328	CCCTGCGGACACTGGTGTGCTGTGTCGTGAACACTCTGCGGGGCATCCAGGCCCTCATGGCAC CC	387
D	307	CCCTGCGGACACTGGTGTGCTGTGTCGTGAACACTCTGCGGGGCATCCAGGCCCTCATGGCAC CC	366
OY	388	CGCGGCTGCCCTGCCACCTGTGTGCTTACGCCACCTGTGACCCCCAGTGTGCTGACACT T	447
D	367	CGCGGCTGCCCTGCCACCTGTGTGCTTACGCCACCTGTGACCCCCAGTGTGCTGACACT T	426
OY	448	ACTGGCAAGCTCGGAGCGCTGCCCTTCCTTTCAGCCTCCATGGCCGACGCTCGGAGGAC CTAG	507
D	427	ACTGGCAAGCTCGGAGCGCTGCCCTTCCTTTCAGCCTCCATGGCCGACGCTCGGAGGAC CTAG	486
OY	508	CCACATTGAGGGCTGAGTCAAGGCTCCCCAACCTTGGCAGACGAGGGGCCACAGGCG CG	567
D	487	CCACATTGAGGGCTGAGTCAAGGCTCCCCAACCTTGGCAGACGAGGGGCCACAGGCG CG	546
OY	568	TAGCATGCGGGGGAGCAGACGCCACCTGTGGTGTGCTTGTGACCTGCTGGGCCACG CACTGG	627
D	547	TAGCATGCGGGGGAGCAGACGCCACCTGTGGTGTGCTTGTGACCTGCTGGGCCACG CACTGG	606
OY	628	CTGTCTACTGAGCANVTGGGCTTGAAGGGCGCTTTTTGAGGATAATTGACACTCTATGTAT GA	687
D	607	CTGTCTACTGAGCANVTGGGCTTGAAGGGCGCTTTTTGAGGATAATTGACACTCTATGTAT GA	666
OY	688	CAATGAACCTTTGGGACACAGCCCTCTGAGGGCCCTCAAMACAGGCCCTGAGATGGGCC GGG	747
D	667	CAATGAACCTTTGGGACACAGCCCTCTGAGGGCCCTCAAMACAGMGCTGAGATGGGCC GGG	726
OY	748	CAAGGAGAGACTCCGAGACTGTGACAGAGCGCGGAATTGACTTACTCATGGA	798
D	727	CAGGAGGAGAACTCCGAGACTGTGACAGAGGTGAATTGACTTACTCATGGA	777

RESULT_11
LOCUS BgB29426 846 bp mRNA linear EST 22-MAY-2001

DEFINITION 6027637J.F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4899146 5' ,
RNA sequence.

ACCESSION BgB29426
VERSION BgB29426.1 GI:14177013
KEYWORDS EST.

SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
*Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 846)*
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strusberg, Ph.D.
Email: rstrsb@remail.nhl.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Lung Hong/Rubin Laboratory
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.jnl.gov>

[illegible]

Db 662 ACATGAACCTTGGGCGACAGCCTCTGTAGGGCCCTCAAAAACAGCCCTGTAGAGATGGCCG 721

Qy 746 GG--CAAGAGAGAAAGCTCCGGAGCTGTAGAGAGGCGGAATTGGACTACCTCATGTATGTG- 802

Db 722 GGACAGAGAGAAAGCTCCGGAGCTGTAGAGAGGCTGTAGACTACCTCATGTATGTGAGT 781

Qy 803 --CTGGTGGGACACAGGCACT--GGAGCAGACCGCGGGGCGGAGGCGCTGAGCCCTGT 858

Db 782 GCTGTGTGGGACACAGGCACTGTAGAGCGAACCAGCGGGGCGGAGGCGCTGAGACATCGT 841

RESULT 12

LOCUS CD364905/c 747 bp mRNA linear EST 29-MAY-2003

DEFINITION UI-H-Ft2-bjn-c-13-0-UI.s1 NCI_CGAP_Ft2 Homo sapiens CDNA clone

ACCESSION CD364905

VERSION CD364905.1 GI:31148995

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 747)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Contact: Robert Strausberg, Ph.D. Unpublished

COMMENT Email: cgaps-remail.nih.gov Tissue Procurement: Dr. Gary W. Hunninghake, U of I CDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/cgap.html> Seq primer: M13 FORWARD

FEATURES

source Location/Qualifiers

1..747

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-H-Ft2-bjn-c-13-0-UI"

/tissue_type="Aveolar Macrophage"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies)"

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker. Site_1: EcoR I; Site_2: Not I. NCI_CGAP_Ft2 is a subcloned cDNA library constructed from a pool of 81 RNA samples from Alveolar Macrophages challenged with different treatments. The library was subtracted according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. The tissue was provided by Dr. Gary W. Hunninghake of the University of Iowa. TAG_LIB=UI-H-Ft2 TAG_TISSUE=Human Lung Aveolar Macrophage TAG_SEQ=GGCGACGCG"

BASE COUNT 146 a 229 c 204 g 166 t 2 others

ORIGIN

Query Match 61.88; Score 737; DB 14; Length 747;

Best Local Similarity 99.78; Pred. NO. 4.1e-162;

Matches 737; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 454 AAGCTGAGAGCTCCCTTTCAGCTCATGGCCAGGCTCTGTAGAGACCTCAGCCACAT 513

Db 747 AAGCTGAGAGCTCCCTTTCAGCTCATGGCCAGGCTCTGTAGAGACCTCAGCCACAN 688

Qy 514 TGAGGGCTGAGTGAAGCTCCCAACCTTGGAGAGAGAGGCGCCAGGCGCTAGCAT 573

Db 687 TGAGGGCTGAGTGAAGCTCCCAACCTTGGAGAGAGAGGCGCCAGGCGCTAGCAT 628

Qy 574 CGGGGAGACAGGCGCCAGCTGTGGTGCCTTGGACCTGTGGGCCAGGCACTGGCTGTCT 633

Db 627 CGGGGAGACAGGCGCCAGCTGTGGTGCCTTGGACCTGTGGGCCAGGCACTGGCTGTCT 568

Qy 634 ACTGAGCATGGGCTTGAAGGCTGTGGAGATTTGACACCTCTATGTATGACATGA 693

Db 567 ACTGAGCATGGGCTTGAAGGCTGTGGAGATTTGACACCTCTATGTATGACATGA 508

Qy 694 ACTTGGGACAGGCTGTGAAGGCTGTGAAGGCTGTGAAGGCTGTGAAGGCTGTGAAGG 753

Db 507 ACTTGGGACAGGCTGTGAAGGCTGTGAAGGCTGTGAAGGCTGTGAAGGCTGTGAAGG 448

Qy 754 GGAAGCTCCGAGCTGTGAAGGCTGTGAAGGCTGTGAAGGCTGTGAAGGCTGTGAAGG 813

Db 447 GGAAGCTCCGAGCTGTGAAGGCTGTGAAGGCTGTGAAGGCTGTGAAGGCTGTGAAGG 388

Qy 814 ACAGGACATGTAGAGGAGCGCGGGGCGGAGGCGCTGAGCCCTGTCTGTGAAGTGTCT 873

Db 387 ACAGGACATGTAGAGGAGCGCGGGGCGGAGGCGCTGAGCCCTGTCTGTGAAGTGTCT 328

Qy 874 TGGTATGTGAAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCT 933

Db 327 TGGTATGTGAAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCT 268

Qy 934 CCTAGTACAGAGAAACAGGCTTGGGCGCACTTGGAGAGACAGATCTAGTCTGGGCA 993

Db 267 CCTAGTACAGAGAAACAGGCTTGGGCGCACTTGGAGAGACAGATCTAGTCTGGGCA 208

Qy 994 CTTACATCCCT 1053

Db 207 CTTACATCCCT 148

Qy 1054 ATGGAATGACAGGCTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1113

Db 147 ATGGAATGACAGGCTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 88

Qy 1114 GTTCCTTCTGTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1173

Db 87 GTTCCTTCTGTGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 28

Qy 1174 GGAATCTCAAAAAA 1192

Db 27 GGAATCTCAAAAAA 9

RESULT 13

LOCUS BG402912 1037 bp mRNA linear EST 12-MAR-2001

DEFINITION 602418715F1 NIH_MGC_93 Homo sapiens CDNA clone IMAGE:452595 5,

ACCESSION BG402912

VERSION BG402912.1 GI:13296360

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1037)

AUTHORS NIH-MGC <http://mhc.ncbi.nlm.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov> Plate: LRAM10432 row: m column: 04 High quality sequence stop: 739.


```

Db      ||| 61 TGGAGAGCTGACCAATGAGGCTGATGACGCTCCCAACCTTGGAGAGAGG 120
Oy      ||| 555 GGCACACAGGCGGTAGCATCGGGGAGAGAGCCCGAGGCTGGGTGCTTGACCTGCTGG 614
Db      ||| 121 GGCACACAGGCGGTAGCATCGGGGAGAGAGAGCCCGAGGCTGGGTGCTTGACCTGCTGG 180
Oy      ||| 615 GGCACACAGGCGGTAGCATCGGGGAGAGAGAGCCCGAGGCTGGGTGCTTGAGATATTGACA 674
Db      ||| 181 GCCACGCGCATGCGCTGCTACTGACGATGGGCTTGAGGCGCTGTTTGAGATATTGACA 240
Oy      ||| 675 CCTATGTATGACATGAACTTTGGGACGAGCGCTGAGGCGCTCAACAGGCGCTG 734
Db      ||| 241 CCTATGTATGACATGAACTTTGGGACGAGCGCTGAGGCGCTCAACAGGCGCTG 300
Oy      ||| 735 AGGATGGGCGGGGCAAGAGAGAGCTCCGAGCTGAGAGAGCGCAATTTGACATCTCA 794
Db      ||| 301 AGGATGGGCGGGGCAAGAGAGAGCTCCGAGCTGAGAGAGCGCAATTTGACATCTCA 360
Oy      ||| 795 TGGATGCTGCTGGGCGACACAGGCACTGGAGCGACCGCGGGCGGCGCTGAGCC 854
Db      ||| 361 TGGATGCTGCTGGGCGACACAGGCACTGGAGCGACCGCGGGCGGCGCTGAGCC 420
Oy      ||| 855 TCGTCTGGAATGTTGCTGCTGATCTGTAAGCTGAGCTGCTGCTGAGCAACTGCTC 914
Db      ||| 421 TCGTCTGGAATGTTGCTGCTGATCTGTAAGCTGAGCTGCTGCTGAGCAACTGCTC 480
Oy      ||| 915 GAAAAGACACAGCTGGCTCCCTAGTACAGAGAGAGGCGCTTGGGCACTTTGAGAGAG 974
Db      ||| 481 GAAAAGACACAGCTGGCTCCCTAGTACAGAGAGAGGCGCTTGGGCACTTTGAGAGAG 540
Oy      ||| 975 AGAATCTAGTCTGGGCGAATTCACATCGGCTCCTGCTGACAGGCGCTGGAGGGGAGC 1034
Db      ||| 541 AGAATCTAGTCTGGGCGAATTCACATCGGCTCCTGCTGACAGGCGCTGGAGGGGAGC 600
Oy      ||| 1035 CTGAATTAACCCCTAGTATGATGATGACAGAGGCTGCTGGGAGATGATTCCTCGGCC 1094
Db      ||| 601 CTGAATTAACCCCTAGTATGATGATGACAGAGGCTGCTGGGAGATGATTCCTCGGCC 659
Oy      ||| 1095 TGGGCTCATGCTTGGGCTGCTTCTGCTGATGATGAGGAGAGAGAGCCCAATCAGATTTT 1154
Db      ||| 660 TGGGCTCATGCTTGGGCTGCTTCTGCTGATGATGAGGAGAGAGAGCCCAATCAGATTTT 719
Oy      ||| 1155 CAATTAAGCAGCTGCGGAAATCTCAAAAAAAA 1192
Db      ||| 720 CAATTAAGCAGCTGCGGAAATCTCAAAAAAAA 756

```

RESULT 15
BU957431
LOCUS BU957431 949 bp mRNA linear EST 21-OCT-2002
DEFINITION AGENCOURT_10621789 NIH_MGC_107 Homo sapiens cDNA clone
IMAGE: 6731456 5', mRNA sequence.
ACCESSION BU957431
VERSION BU957431.1 GI:24187003
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: c9apds@email.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov

```

FEATURES
source
Plate: LNCM3057 row: 0 column: 07
High quality sequence stop: 670.
Location/Qualifiers
1..949
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6731456"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_idb="NIH_MGC_107"
/note="Organ: breast; Vector: pORF7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 221 a 264 c 289 g 174 t 1 others
ORIGIN
Query Match 61.0%; Score 727; DB 13; Length 949;
Best Local Similarity 94.7%; Pred. No. 9.5e-160;
Matches 809; Conservative 0; Mismatches 31; Indels 14; Gaps 5;
Oy 353 AACACTGCGGCGCATCAGGCGCTGCATGACCGCGCGGCTGCCCTGACCTGTGCT 412
Db 2 AACACTGCGGCGCATCAGGCGCTGCATGACCGCGCGGCTGCCCTGACCTGTGCT 61
Oy 413 AGCCACCTGACGCGCCGAGTGGCTGACAACTTACGCACTGAGAGCTGGAGCGCTT 472
Db 62 AGCCACCTGACGCGCCGAGTGGCTGACAACTTACGCACTGAGAGCTGGAGCGCTT 121
Oy 473 TCAGCTTCATAGGCGCAGGCTCCTGAGAGAGCTCAGCAGCATTTGAGGCGTGAAGCT 532
Db 122 TCAGCTTCATAGGCGCAGGCTCCTGAGAGAGCTCAGCAGCATTTGAGGCGTGAAGCT 181
Oy 533 CCCCAACCTTGGCAGAGAGAGGCGCCACAGCGCGTACAGTGGGAGAGAGCGCCAGC 592
Db 182 CCCCAACCTTGGCAGAGAGAGGCGCCACAGCGCGTACAGTGGGAGAGAGCGCCAGC 241
Oy 593 CTGGGTGCTTGGACCTGCTGGGCGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 652
Db 242 CTGGGTGCTTGGACCTGCTGGGCGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 301
Oy 653 GGCCTGTTGAGATTTGACACCTCTATGATGATGACATGAACTTTGGGACAGCCTCT 712
Db 302 GGCCTGTTGAGATTTGACACCTCTATGATGATGACATGAACTTTGGGACAGCCTCT 361
Oy 713 GAGGCGCTCAACAGAGCGCTGAGAGAGGCGCGGCAAGAGAGAGAGCTCCGAGAGCT 772
Db 362 GAGGCGCTCAACAGAGCGCTGAGAGAGGCGCGGCAAGAGAGAGAGCTCCGAGAGCT 421
Oy 773 GAGGCGAATTTGAGCTACCTCATGATGATGCTGCTGGGCGACAGAGAGAGAGAGCG 832
Db 422 GAGGCGAATTTGAGCTACCTCATGATGATGCTGCTGGGCGACAGAGAGAGAGAGCG 481
Oy 833 CCGGCGCAGGCGCGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 892
Db 482 CCGGCGCAGGCGCGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541
Oy 893 GCTGGCGTGAACAAGTCTCTCGAAAGAGACAGAGCTGCTCCCTAGTACAGAGAGAG 952
Db 542 GCTGGCGTGAACAAGTCTCTCGAAAGAGACAGAGCTGCTCCCTAGTACAGAGAGAG 601
Oy 953 GCTGGCGCAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1012
Db 602 GCTGGCGCAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
Oy 1013 TCTCAGGCGTGGCAGGAGGAGCTGGAATTACCCCTAGTATGATGATGATGATGATG 1070
Db 662 TCTCAGGCGTGGCAGGAGGAGCTGGAATTACCCCTAGTATGATGATGATGATGATG 721

```



```

Qy      1071 GGTGGGACTGAATTCCTGGCCC--TGGGTCATAGCTTGGGCTGTCTCTCT---G 1125
Db      722  GTGGGGGACTGAATTCCTGGNCCCTGGGGGTCATACCTTGGGCTGTCTCTCTGA 781
Qy      1126 ATACGGGAGAGAGACCCCA---ATCAGATTTTCAAAATTAAGCC---AGTCCTGGGAAA 1178
Db      782  AAAGGGGAGAGAGACCCCAATCAGAAATTTTCAAAATTAAGCCCAAGTCTCTGGGGAAA 841
Qy      1179 TCTCAAAAAAAAAA 1192
Db      842  TCTCAAAAAAAAAA 855

```

Search completed: August 25, 2003, 16:45:27
 job time : 2467 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 13:45:31 ; Search time 97 Seconds
(without alignments)
5424.007 Million cell updates/sec

Title: US-09-701-675A-8

Perfect score: 1192
Sequence: 1 gtagccggagtcagactg.....gggaattctcaaaaaaaaaa 1192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 segs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/lna/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/lna/OCUTS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	45.6	3.8	1056	US-09-252-991A-12106	Sequence 12106, A
C 2	45.6	3.8	3132	US-09-252-991A-12314	Sequence 12314, A
C 3	44.6	3.7	2639	US-09-620-312D-773	Sequence 773, App
C 4	43.6	3.7	702	US-09-252-991A-8450	Sequence 8450, Ap
C 5	43.6	3.7	1278	US-09-252-991A-8324	Sequence 8324, Ap
C 6	43.6	3.7	1362	US-09-252-991A-8483	Sequence 8483, Ap
C 7	43.2	3.6	636	US-09-252-991A-8356	Sequence 8356, Ap
C 8	39.8	3.3	1314	US-09-252-991A-3638	Sequence 3638, Ap
C 9	39.8	3.3	2622	US-09-252-991A-15001	Sequence 15001, A
C 10	38.8	3.3	1095	US-09-252-991A-14880	Sequence 14880, A
C 11	38.8	3.3	1317	US-09-252-991A-14748	Sequence 14748, A
C 12	38.4	3.2	2109	US-09-370-838-153	Sequence 153, App
C 13	38.4	3.2	4822	US-09-484-970B-106	Sequence 106, App
C 14	38.2	3.2	4185	US-08-964-268-1	Sequence 1, App11
C 15	37.6	3.2	4185	US-09-105-254-1	Sequence 1, App11
C 16	37.6	3.2	4847	US-09-045-632-47	Sequence 47, App1
C 17	37.6	3.2	4847	US-09-045-632-48	Sequence 48, App1
C 18	37.4	3.1	846	US-09-252-991A-3196	Sequence 3196, Ap
C 19	37.4	3.1	2539	US-09-000-016-3	Sequence 3, App11
C 20	37.4	3.1	2539	US-09-514-340-3	Sequence 3, App11
C 21	37.4	3.1	2809	US-09-000-016-1	Sequence 1, App11
C 22	37.4	3.1	2809	US-09-514-340-1	Sequence 1, App11
C 23	37.4	3.1	1128	US-09-622-439-1	Sequence 1, App11
C 24	37.2	3.1	344	US-09-060-756-718	Sequence 718, App
C 25	37	3.1	344	US-09-670-314-718	Sequence 718, App
C 26	37	3.1	780	US-09-252-991A-16418	Sequence 16418, A
C 27	36.8	3.1	780	US-09-252-991A-16418	Sequence 16418, A

28	36.8	3.1	960	US-09-252-991A-16064	Sequence 16064, A
29	36.8	3.1	2430	US-09-252-991A-16171	Sequence 16171, A
C 30	36.8	3.1	3675	US-09-252-991A-16309	Sequence 16309, A
C 31	36.6	3.1	468	US-09-252-991A-16322	Sequence 16322, A
C 32	36.6	3.1	1289	US-09-252-991A-16434	Sequence 16434, A
C 33	36.6	3.1	1299	US-09-252-991A-16434	Sequence 16434, A
C 34	36.6	3.1	1566	US-09-252-991A-16161	Sequence 16161, A
C 35	36.6	3.1	1662	US-09-252-991A-16050	Sequence 16050, A
C 36	36.6	3.1	2028	US-09-252-991A-10455	Sequence 10455, A
C 37	36.6	3.1	2094	US-09-252-991A-10712	Sequence 10712, A
C 38	36.6	3.1	2338	US-08-425-069-1	Sequence 1, App11
C 39	36.6	3.1	2338	US-08-317-844B-1	Sequence 1, App11
C 40	36.6	3.1	2681	US-08-928-213B-7	Sequence 7, App11
C 41	36.6	3.1	4403765	US-09-103-840A-2	Sequence 2, App11
C 42	36.6	3.1	4411529	US-09-103-840A-1	Sequence 1, App11
C 43	36.4	3.1	1053	US-09-904-615-22	Sequence 22, App1
C 44	36.4	3.1	1712	US-09-620-312D-1101	Sequence 1101, Ap
C 45	36	3.0	4600	US-09-702-705-1797	Sequence 1797, Ap

ALIGNMENTS

```
RESULT 1
US-09-252-991A-12106/C
; Sequence 12106, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12106
; LENGTH: 1056
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-12106

Query Match      3.8%   Score 45.6; DB 4; Length 1056;
Best Local Similarity 55.6%; Pred. No. 0.02; 84; Indels 3; Gaps 1;
Matches 109; Conservative 0; Mismatches 84; Indels 3; Gaps 1;

QY      220 TCCTGGCAGCGAGCGGTGGACAGCCCGCGCGTGGCC---TCTAGCTCCCTTT 276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1036 TCCGGGTCTATGCCCGCATGACAGGAAATTCCTGGCGGTGACGAAACCGGCTCTCGG 977

QY      277 TGACCTCTAGTCTCAGCTCCACAGCCTTCAGCAGAGTAGAGCGGAGCTGGGCA 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      976 AGAGCGCCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 917

QY      337 CCGTGGCGTGGTGGTGAACACTCTGGGGGCGATCCAGGGGTGCATGGACCCGGCGTCG 396
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      916 GCTGGTCTGCTGCTCAGCTGGTCAAGCGGCGAGCGGCTGCGGCGGCTGCGGGGTAT 857

QY      397 CCTGCACCTGTGCTT 412
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      856 CGATCCAGCGCGGGGCT 841

RESULT 2
US-09-252-991A-12314
; Sequence 12314, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
```

;; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 12314
;; LENGTH: 3132
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12314

Query Match 3.8%; Score 45.6; DB 4; Length 3132;
Best Local Similarity 55.6%; Pred. No. 0.029;
Matches 109; Conservative 0; Mismatches 84; Indels 3; Gaps 1;

QY 220 TCCCTGACCGACGCGGTGGACAGGACACCCCGCGCTGGCC--TCTAGCTCCCTCTT 276
DB 2088 TCCGGGTATGCGCCATCGACAGGAAATCTCTGGCCGGTGACGAACCGGCGTCTCGG 2147
QY 277 TGACCTCTAGTGTCAAGCTCCACACAGCCTGAGCAGAGTGAAGCCGAGCTGCGCA 336
DB 2148 AGAGCAGCCACAGCCACGACGAGGACATCTCCACCGTTTCCGACCGCGGACACAGT 2207
QY 337 CCTGTGCTGTGCTGTAACACTCTGCGGCGCATCCAGCGCTTCATGACACCGCGCTGC 396
DB 2208 GCTGTGCTGTGCTGCTGCGGCTGCGACGCGGCGAGCTCGGCGTCCGCGGCTCGCGGTAT 2267
QY 397 CCGTGCACCTGTGCCT 412
DB 2268 GATCCAGCGCGGCGCT 2283

RESULT 3
US-09-620-312D-773
; Sequence 773, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundun
; APPLICANT: Wang, Duntui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanec, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: Polyptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pf_fl_genes Version 1.0
; SEQ ID NO 773
; LENGTH: 2629
; TYPE: DNA
; ORGANISM: Homo sapiens

;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (73)..(798)
US-09-620-312D-773

Query Match 3.7%; Score 44.6; DB 4; Length 2629;
Best Local Similarity 50.7%; Pred. No. 0.045;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 269 TCCCTTTGACCTCTAGTGTCAAGCTCCACACAGCCTGACAGAGTGAAGCCGAG 328
DB 178 TCGCTCTGAGATCTCTGTGTGAAGTTGACAGCTTTGCCATGCTTGTGTGAGCCCAAC 237
QY 329 CCGGACCTGTGCTGTGCTGTGAACACTCTGCGCGCATCCAGGCGTCCATGGCACCC 388
DB 238 CTGTGCGCTGAGTCTCTATTGCCAACAGGTCGCGGAGATCCAAAGAGATGACGAG 297
QY 389 GCGGTGCGCTGCGCACCTGTGCTGAGCCGACCTGAGCCGAGTGTGCTGACACTTA 448
DB 298 GATGGAGCTGCGCGACAGTGGCACCCGACGCTGAGAGCGGCGCGCTGACCGCTTG 357
QY 449 CTGGCAAGCTGAGAGCTGCGCTTTCAGCCT 479
DB 358 GTCTCCAGGAGATCTGTGCGTGCACAGCT 388

RESULT 4
US-09-252-991A-8450/C
; Sequence 8450, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8450
; LENGTH: 702
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8450

Query Match 3.7%; Score 43.6; DB 4; Length 702;
Best Local Similarity 43.7%; Pred. No. 0.057;
Matches 193; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

QY 152 GGTCTGAAGCGGAACGGGAGAGAGAGAGAGAACTCTGAGTGCATCTCTGG 211
DB 661 GGTCAAGCGCGCGGTGCGCGCGAGGTCACACAGCGGATGCCCCGGGCTCTGCGGAG 602
QY 212 TGGCTAATCTGTGCGACAGCGGCTGACACAGCACCCCGCGGCTGAGCTTAC 271
DB 601 CAGCAGGCTGCGGTGCGGAGTCCACCCGTCGCGGAGGAGTCCAGAGGCGAAGGCGC 542
QY 272 CTCTTTGACCTCTAGTGTCAAGCTCCACACAGCTGAGAGTGAAGCCGAGACTG 331
DB 541 GCGGCGGTCTCTCTTCACTTCCAGGCGCATGCGCGCGCGCGGCTGCGGCTGCTG 482
QY 332 CGGACCTGTGTGTGTGTGAACACTCTGCGGCGCATCCAGGCGTCCATGGCACCGCG 391
DB 481 GCGCGCGTGCAGCAGCGGTGCGCGCTGCTGCTGCGGCGGCTGAGGCGGCGGCTGCTG 422
QY 392 GCTGCGCTGCGACCTGTGCTGAGCCACTGACGCGCCCGCATGTGTGCTGACAACTTGTG 451
DB 421 GCGCAGCAACCAAGTCGCGGCGCTGCGGCTGCGGCTGCGGCTGCGGCGGAGACCAT 362
QY 452 GCAAGCTCGAGCGCTGCGCTTTCAGGCTTCATGCGCAGCCCTCTGTGAGAGCACTGACGAC 511

```
Db      361 GCGGCGCTGAGACCCAGAGGCTCCAGGCGCGCGCGGCGAGAGAGCAAGACGCCGAG 302
QY      512 ATTGAGGCGCTGAGTACGAGCTCCCAACCTTGGACAGAGAGGCGGCGACCGGAGG 571
Db      301 GGGGCGCGCGAGAGAGGCGGCGGCTCGACGAGCTGCTCCAGGCGCGGCGAGCAGGTTAG 242
QY      572 ATCGGGGAGAGCAGCGCCAGCC 593
Db      241 GCTCGGCTGGCGCGCGCAGGGC 220
```

RESULT 5

```
US-09-252-991A-8324
; Sequence 8324, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8324
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8324
```

```
Query Match      3.7%; Score 43.6; DB 4; Length 1278;
Best Local Similarity 43.7%; Pred. No. 0.069;
```

```
Matches 193; Conservative 0; Mismatches 249; Indels 0; Gaps 0;
```

```
QY      152 GGTCTGAGCGGAAACGGAGAGAGAGAGAGAAACCTTGGCAGTGCAGTCTCTGG 211
Db      168 GGTGACCGCGCGCTCCGCGGAGGGTTCACAGAGCGGAGTCCCGCGCTCTCGGCGAG 227
QY      212 TGGCTGATCTTGGCCAGCAGGAGTGGACAGAGCAGCCCGGCGCTTGGCTTACTTC 271
Db      228 CACAGAGCTGCGCGCTGCGGCAATGCACCGTCCGCGGCGGCACTCGAGGCGGAAAGCGCC 287
QY      272 CTTTGAAGCTCTGAGTGTCAAGTCTCAACAGCAGCAGCAGTGCAGAGAGTGGAGCGACTG 331
Db      288 GCGCGGCTCTCTTACTTCACTTCAGGGGCAATCGGCGCGCGCGGCGGCACTCCAGGCTGCTG 347
QY      332 CGGACCTGAGTGTGCTGTGAACACTTGGCGGCGCATCGAGCGCTCATGGACCGCGG 391
Db      348 GCGCGCTGAGAGAGGGTTCGCGCTTGGTCCGCGGTTCCAGGCGGAGGCGGCTGCGGTC 407
QY      392 GCTGCGCTGCGCACTGTGCTAGCCCACTGCAAGCGCCCAAGTGTGCTGACAACTTACTG 451
Db      408 GCGCAGCAACCAAGTCCGCGGCTCGCGGCTCGGCGGCAATTCGCGCGGCGAGACCAT 467
QY      452 GCNAGCTGAGAGCTCCCTTTCAGGCTCCATGCGCAGGCGCTCTGGAGAGCTCAGCCAG 511
Db      468 GCGCGCTGAGACCCAGAGGCTCCAGGCGGCGCGCGGCGGAGAGAGCAAGACGCCGAG 527
QY      512 ATTGAGGCGCTGAGTACGCTCCCAACCTTGGCAGAGAGGCGGCGCAGGCGCTAGC 571
Db      528 GCGCGCGCGAGAGACCGCGGCGGCTCGACGCGCTCTCCAGGCGCGGAGCAGCAGGTTAG 587
QY      572 ATCGGGGAGAGCGCCAGCC 593
Db      588 GCTCGGCTGGCGCGCGCAGGGC 609
```

```
RESULT 6
US-09-252-991A-8483/c
```

```
; Sequence 8483, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8483
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8483
```

```
Query Match      3.7%; Score 43.6; DB 4; Length 1362;
Best Local Similarity 43.7%; Pred. No. 0.071;
```

```
Matches 193; Conservative 0; Mismatches 249; Indels 0; Gaps 0;
```

```
QY      152 GGTCTGAGCGGAAACGGAGAGAGAGAGAGAAACCTTGGCAGTGCAGTCTCTGG 211
Db      1029 GGTGACCGCGCGCTCCGCGGAGGGTTCACAGAGCGGAGTCCCGCGCTCTCGGCGAG 970
QY      212 TGGCTGATCTTGGCCAGCAGGAGTGGACAGAGCAGCCCGGCGCTTGGCTTACTTC 271
Db      969 CAGCAGGCTGCGCGCTGCGGCAATGCACCGTCCGCGGCGGCACTCGAGGCGGAAAGCGCC 910
QY      272 CTTTGAAGCTCTGAGTGTCAAGTCTCAACAGCAGCAGCAGTGCAGAGAGTGGAGCGACTG 331
Db      909 GCGCGGCTCTCTTACTTCACTTCAGGGGCAATCGGCGCGCGGCGGCACTCCAGGCTGCTG 850
QY      332 CGGACCTGAGTGTGCTGTGAACACTTGGCGGCGCATCGAGCGCTCATGGACCGCGG 391
Db      849 GCGCGCTGAGAGAGGGTTCGCGCTTGGTCCGCGGTTCCAGGCGGAGCGGCTGCGGTC 790
QY      392 GCTGCGCTGCGCACTGTGCTAGCCCACTGCGGCGCATCGAGCGCCCAAGTGTGCTGACAACTTACTG 451
Db      789 GCGCAGCAACCAAGTCCGCGGCTCGCGGCTCGGCGGCAATTCGCGCGGCGAGACCAT 730
QY      452 GCNAGCTGAGAGCTCCCTTTCAGGCTCCATGCGCAGGCGCTCTGGAGAGCTCAGCCAG 511
Db      729 GCGCGCTGAGACCCAGAGGCTCCAGGCGGCGCGGCGGCGGAGAGAGCAAGACGCCGAG 670
QY      512 ATTGAGGCGCTGAGTACGCTCCCAACCTTGGCAGAGAGGCGGCGCAGGCGCTAGC 571
Db      669 GCGCGCGCGAGAGACCGCGGCGGCTCGACGCGCTCTCCAGGCGCGGAGCAGCAGGTTAG 610
QY      572 ATCGGGGAGAGCGCCAGCC 593
Db      609 GCTCGGCTGGCGCGCGCAGGGC 588
```

RESULT 7

```
US-09-252-991A-8356
; Sequence 8356, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
```


.,Tue Aug 26 13:49:14 2003

us-09-701-675a-8.rni

Page 7

Search completed: August 25, 2003, 16:47:17
Job time : 104 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 12:16:56 ; Search time 371 Seconds
(without alignments)
8673.127 Million cell updates/sec

Title: US-09-701-675a-8

Perfect score: 1192
Sequence: 1 gtagcgcgagtcagcagctg.....gggaatctcaaaaaaaaa 1192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_19jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1192	100.0	1192	21	AAZ29482	DNA encoding human
2	1188.8	99.7	1219	22	AAH64895	Human secreted pro
3	1183.4	99.3	2113	22	AAK94674	Human full-length
4	1167	97.9	1181	22	AAH99491	Human protein enco
5	1154.4	96.8	1222	21	AACT6627	Human ORF2182
6	1151.2	96.6	1182	20	AAK97688	Extended human sec
7	1061.2	89.0	1078	20	AAZ19894	Human adult blood
8	889.4	74.6	1216	23	AAZ78542	DNA encoding novel

9	777.8	65.3	5746	24	ABK81819	DNA representing 1
10	707.8	59.4	711	20	AAK36177	DNA encoding a p16
11	667.4	56.0	5869	24	AAI49936	Human molecule for
12	637.2	53.5	696	22	AAK91707	Human CDNA 5'-end
13	637.2	53.5	696	22	AAK93731	Human CDNA clone r
14	447	37.5	447	21	AAZ29502	Human CDNA clone r
15	433.8	36.4	450	20	AAV86113	EST clone G21. Ho
16	293.4	24.6	301	20	AAH41286	Human secreted pro
17	237.6	19.9	363	24	AAZ62144	Porcine muscular s
18	232	19.5	356	24	ABZ11398	Human polynucleoti
19	189	15.9	189	21	AAZ29488	Incyle clone 13520
20	188	15.8	190	21	AAZ29504	Incyle clone 30157
21	120.8	10.1	165	24	AAZ69457	Novel murine polyn
22	60	5.0	60	24	ABK39710	Human spliced tran
23	52.2	4.4	65	24	ABK55795	Mouse spliced tran
24	49.2	4.1	1253	24	ABK97063	Human NOV34a, CDNA.
25	44.6	3.7	862	22	AAH06538	Human CDNA clone (
26	44.6	3.7	1314	24	ABL49854	Human CHD protein
27	44.6	3.7	2082	24	AAI41498	Human haematopoiet
28	44.6	3.7	2447	24	AAZ34182	Human CDNA coexpre
29	44.6	3.7	2629	22	AAI58881	Human polynucleoti
30	44.6	3.7	2629	22	AAI60667	Human polynucleoti
31	44.2	3.7	4356	14	AAQ37543	Cardiac adenyllyl c
32	43	3.6	396	22	AAE94876	Human adenyllyl c
33	43	3.6	396	22	ABT03243	Human ovarian carc
34	43	3.6	396	24	ABL48926	Human ovarian carc
35	43	3.6	2038	21	AAE18134	Lung cancer associ
36	43	3.6	2038	22	AAH33113	Human colon cancer
37	42.6	3.6	4356	16	AAQ95540	Cardiac adenyllyl c
38	42.4	3.6	1387	13	AAQ20688	GDF-1a cDNA sequen
39	42.4	3.6	1387	13	AAZ49675	Mouse Gli3 cell 1
40	42.2	3.5	1428	25	AAZ53024	Streptomyces albol
41	42.2	3.5	2082	24	AAI41497	Mouse haematopoiet
42	42.2	3.5	122748	24	ABT10719	Human breast cance
43	41.8	3.5	1413	25	AAZ53019	Streptomyces timos
44	41.4	3.5	2949	25	ABZ71137	S. mutrayaensis A
45	41.4	3.5	36321	25	ABZ71131	Streptomyces muray

ALIGNMENTS

RESULT 1	AAZ29482	standard; DNA; 1192 BP.
ID	AAZ29482	
XX	AAZ29482;	
AC	14-MAR-2000	(first entry)
XX		
DT		
DE		DNA encoding human cell cycle regulation protein-3.
XX		
KW	CEBP-3; cell cycle regulation protein-3; tissue expression; cancer;	
KW	inflammation; phlebotomy; gene therapy; disease diagnosis;	
KW	disease monitoring; chromosomal mapping; cell proliferation;	
KW	cell proliferative disorder; ds.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	140..850
FT		/*tag= a
FT		/product= "cell cycle regulation protein-3"
PN	MO9964593-A2.	
XX		
PD	16-DEC-1999.	
XX		
PF	08-JUN-1999;	99WO-US12906.
XX		
PR	08-JUN-1999;	98US-0088695.
XX		
PA	(INCYTE-)	INCYTE PHARM INC.

XX Bandman O, Lal P, Tang YT, Corley NC, Guejler KJ, Baughn ME;
PI Patterson C;
XX
XX WPI: 2000-105887/09.
DR P-PSDB: AA44362.
XX
XX Novel regulatory proteins, for diagnosis, treatment and prevention of
PT cell proliferative and immune system diseases
XX
XX
PS Claim 7; Page 75; 88pp; English.
CC The present sequence encodes cell cycle regulation protein-3 (CECRP-3).
CC Tissue expression is seen in reproductive (0.302), nervous (0.163) and
CC cardiovascular (0.186) tissues. The cDNA is cloned into Bluescript
CC vector. Diseases associated with tissue expression of CECRP-3 are cancer,
CC inflammation and other diseases. This is used in the recombinant
CC production of CECRP-3, in gene therapy (including expression of antisense
CC sequences, triplex-forming molecules or ribozymes), and as source of
CC probes and primers for diagnosis and monitoring of disease. It is also
CC used for chromosomal mapping and isolation of related sequences. CECRPs
CC are activators of cell proliferation or inhibitors of cellular processes
CC that modulate proliferation. They are used to treat cell proliferative
CC disorders.
XX
XX Sequence 1192 BP; 237 A; 350 C; 375 G; 230 T; 0 other;
SQ
Query Match 100.0%; Score 1192; DB 21; Length 1192;
Best Local Similarity 100.0%; Pred. No. 1,2-233;
Matches 1192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTGAGCCGGAGTCAAGTGGCTGCGGAGCCAGCGGGTTCCGGAGAGACGCCA 60
DB 1 GTGAGCCGGAGTCAAGTGGCTGCGGAGCCAGCGGGTTCCGGAGAGACGCCA 60
OY 61 CAAGGATGTCGCCCGCCGTTTCTGATTGGTGGTGGTGGTGGTGGTGGTGGT 120
DB 61 CAAGGATGTCGCCCGCCGTTTCTGATTGGTGGTGGTGGTGGTGGTGGTGGT 120
OY 121 TTGGCCGCTAGTACCAAGATGCTGAGCAAGGGTCTGAAGCGGAAGCGAGAGAGA 180
DB 121 TTGGCCGCTAGTACCAAGATGCTGAGCAAGGGTCTGAAGCGGAAGCGAGAGAGA 180
OY 181 GGAGAAGGAACCTTGGGAGTGCAGTCTGCTGGTGAATCTCTGGCCAGCGGTGGC 240
DB 181 GGAGAAGGAACCTTGGGAGTGCAGTCTGCTGGTGAATCTCTGGCCAGCGGTGGC 240
OY 241 ACAGGACACCCCGGCGCTGAGCTAGCTCCCTTTGACCTCTGAGTCAAGCTCCA 300
DB 241 ACAGGACACCCCGGCGCTGAGCTAGCTCCCTTTGACCTCTGAGTCAAGCTCCA 300
OY 301 CCACAGCTGACAGAGAGTGAAGCGGAGCTGGGACCTGGTGGTGGTGGTGGTGGT 360
DB 301 CCACAGCTGACAGAGAGTGAAGCGGAGCTGGGACCTGGTGGTGGTGGTGGTGGT 360
OY 361 GGCGGCGATCCAGGGGTCATGGGACCGCGGGTCCCTGGCACCTGGCTAGGCCAGC 420
DB 361 GGCGGCGATCCAGGGGTCATGGGACCGCGGGTCCCTGGCACCTGGCTAGGCCAGC 420
OY 421 TGACAGCCCGCAGTGTGGTGAACAATTACTGGCAAGCTGGAGAGTGGCTTTGAGCCTC 480
DB 421 TGACAGCCCGCAGTGTGGTGAACAATTACTGGCAAGCTGGAGAGTGGCTTTGAGCCTC 480
OY 481 CATGGCCAGCTCTGAGAGACCTGACCAATTTGAGGGGCTGAGTCAAGCTCCCAAC 540
DB 481 CATGGCCAGCTCTGAGAGACCTGACCAATTTGAGGGGCTGAGTCAAGCTCCCAAC 540
OY 541 CTTGGCAGAGAGAGGGGCCACAGGCGCTGAGATCGGGGAGACAGCGCCACCTGGGTGC 600
DB 541 CTTGGCAGAGAGAGGGGCCACAGGCGCTGAGATCGGGGAGACAGCGCCACCTGGGTGC 600
OY 601 CTTGAGACTGTGGGCCAGGCACTGAGTGTCTACTGAGCATGGAGGCTTGAAGGCGCTTT 660
DB 601 CTTGAGACTGTGGGCCAGGCACTGAGTGTCTACTGAGCATGGAGGCTTGAAGGCGCTTT 660

DB 601 CTTGAGACTGTGGGCCAGGCACTGAGTGTCTACTGAGCATGGAGGCTTGAAGGCGCTTT 660
OY 661 TGAGGATATATGACACCTTATGTATGACATGAACTTTGGGACACGAGCTCTGAGGAGCT 720
DB 661 TGAGGATATATGACACCTTATGTATGACATGAACTTTGGGACACGAGCTCTGAGGAGCT 720
OY 721 CAAACAGGCGCTTGGAGATGGGCGGGGCAAGAGAGAACTCCGAGCTGACGAGGCCGA 780
DB 721 CAAACAGGCGCTTGGAGATGGGCGGGGCAAGAGAGAACTCCGAGCTGACGAGGCCGA 780
OY 781 AATGGACTACCTCATGATGT 840
DB 781 AATGGACTACCTCATGATGT 840
OY 841 AGGCGCTGAGACCCCTGCTGAGTGAATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
DB 841 AGGCGCTGAGACCCCTGCTGAGTGAATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
OY 901 GACCAACTGTCTCTGAAAAGACAGAGCTGGCTTCCCTAGTACAGAGAGAGAGGCTTGGGC 960
DB 901 GACCAACTGTCTCTGAAAAGACAGAGCTGGCTTCCCTAGTACAGAGAGAGAGGCTTGGGC 960
OY 961 CACTTTGGAGAGACAGAACTAGTCTGTGGCAACTTCAATCCGCTCTGCTCAAGG 1020
DB 961 CACTTTGGAGAGAGAGAACTAGTCTGTGGCAACTTCAATCCGCTCTGCTCAAGG 1020
OY 1021 CTGGCAGGGGAGACCTGGAATTTACCCCTAGTATGATGATGACAGAGGCTGTGGGACT 1080
DB 1021 CTGGCAGGGGAGACCTGGAATTTACCCCTAGTATGATGATGACAGAGGCTGTGGGACT 1080
OY 1081 GAATTCCTGGCCCTGGGGGTATAGCTTGGGCTGTCTCTGTCTGTCTGTCTGTCTGTCT 1140
DB 1081 GAATTCCTGGCCCTGGGGGTATAGCTTGGGCTGTCTCTGTCTGTCTGTCTGTCTGTCT 1140
OY 1141 CCAATCAGATTTTCAATTAAGCCAGTCTGGGAAATTCAAAAGAGAGAGAGAGAGAG 1192
DB 1141 CCAATCAGATTTTCAATTAAGCCAGTCTGGGAAATTCAAAAGAGAGAGAGAGAGAG 1192

RESULT 2
AAH64895
ID AAH64895 standard; cDNA; 1219 BP.
XX
XX AAH64895;
XX
XX 11-SEP-2001 (first entry)
XX
XX Human secreted protein cDNA, SEQ ID NO: 171.
XX
XX Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;
XX GENSET; ss.
XX
XX Homo sapiens.
XX
XX WO200142451-A2.
XX
XX 14-JUN-2001.
XX
XX 07-DEC-2000; 2000WO-1B01938.
XX
XX 08-DEC-1999; 99US-0169629.
XX PR 06-MAR-2000; 2000US-0187470.
XX
XX (GEST) GENSET.
XX
XX Dumas Milne Edwards J, Bougueleret L, Jobert S;
XX
XX WPI: 2001-367870/38.
XX P-PSDB: AAG89292.
XX
XX Full length GENSET human nucleic acids encoding potentially secreted
XX proteins, useful in gene therapy and vaccination against a variety of
XX diseases, and for diagnosis of those diseases -

XX Claim 7; Page 733-734; 921pp: English.
CC The invention relates to full length GENSET human nucleic acids encoding
CC potentially secreted proteins. The nucleic acids and the polypeptides
CC they encode may be used in the prevention, treatment and diagnosis of
CC diseases associated with inappropriate GENSET gene expression. For
CC example, they be used to treat disorders associated with decreased
CC GENSET gene expression by rectifying mutations or deletions in a
CC patient's genome that affect the activity of GENSET or by supplementing
CC the patients own production of GENSET polypeptides. Conversely,
CC antisense nucleic acid molecules may be administered to down regulate
CC GENSET expression by binding with the cells' own genes and preventing
CC their expression. The sense and antisense nucleic acids may also be
CC used as DNA probes in diagnostic assays to detect and quantitate the
CC presence of similar nucleic acid sequences in samples, and hence to
CC determine which patients may be in need of restorative therapy.
CC The GENSET polypeptides may be used as antigens in the production of
CC antibodies and in assays to identify modulators (agonists and
CC antagonists) of GENSET polypeptide expression and activity. The
CC present sequence is a GENSET nucleic acid of the invention.
XX
SQ Sequence 1219 BP; 249 A; 352 C; 381 G; 237 T; 0 other;

Query Match 99.7%; Score 1188.8; DB 22; Length 1219;

Best Local Similarity 99.8%; Pred. No. 7.8e-293;

Matches 1190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GTGAGCCGAGTCAAGTCTGCGAGCCAGCGCGGTTCCGAGAGACAGCCAA 60
21 GTGAGCGGAGTCAAGTCTGCGAGCCAGCGCGGTTCCGAGAGACAGCCAA 80
61 CAAGCGAGTCTGCGAGCCAGCGCGGTTCCGAGAGTCTGCGAGTCTGCGAG 120
81 CAAGCGAGTCTGCGAGCCAGCGCGGTTCCGAGAGTCTGCGAGTCTGCGAG 140
121 TTGGCGCGTATGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 180
141 TTGGCGCGTATGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 200
181 GGAGAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 240
201 GGAGAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 260
241 ACAGGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 300
261 ACAGGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 320
301 CCAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 360
321 CCAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 380
361 GCGGCGCATGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGAGA 420
381 GCGGCGCATGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGAGA 440
421 TGCAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 480
441 TGCAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 500
481 CATGGCAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 540
501 CATGGCAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 560
541 CTTGGCAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 600
561 CTTGGCAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 620
601 CTTGGCAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 660
621 CTTGGCAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGAGA 680
661 TGAGATATGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 720

Db 681 TGAGATATGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 740
OY 721 CAACAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 780
Db 741 CAACAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 800
OY 781 ATTGAGATATGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGA 840
Db 801 ATTGAGATATGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGA 860
OY 841 AGGCGCTGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 900
Db 861 AGGCGCTGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 920
OY 901 GAGCAAGTCTGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGA 960
Db 921 GAGCAAGTCTGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGA 980
OY 961 CACTTGGAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 1020
Db 981 CACTTGGAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 1040
OY 1021 CTGGCAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 1080
Db 1041 CTGGCAGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 1100
OY 1081 GAATTCCTGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 1140
Db 1101 GAATTCCTGAGAGTCTGAGAGTCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 1160
OY 1141 CCAATCAGATTTTCAATTAAGGCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 1192
Db 1161 CCAATCAGATTTTCAATTAAGGCTGAGCAAGGCTGAGCAAGGAGAGAGAGA 1212
RESULT 3
AAK94674
ID AAK94674 standard; cDNA; 2113 BP.
XX
AC AAK94674;
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 3676.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99UP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI: 2001-52425/58.
DR P-PSDB: AAM93724.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 3676; 1380pp + sequence listing; English.

CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5' and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesising the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX Sequence 2113 BP, 453 A; 561 C; 578 G; 521 T; 0 other;

Query Match 99.3%; Score 1183.4; DB 22; Length 2113;
 Best Local Similarity 99.9%; Pred. No. 2.2e-291;
 Matches 1184; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGAGCCGGAGTCAGACTGCGTCTCGAGACCCAGCGCGGTTCCGGAGACAGCCAA 60
 DB 21 GTGAGCCGGAGTCAGACTGCGTCTCGAGACCCAGCGCGGTTCCGGAGACAGCCAA 80
 QY 61 CAAGGATGTCGCGCGCGCGGTTCCGATGTTGGTGGTCTACTCTTCCTTCGTA 120
 DB 81 CAAGGATGTCGCGCGCGCGGTTCCGATGTTGGTGGTCTACTCTTCCTTCGTA 140
 QY 121 TTGGCCGCTAGTAGACAGATGCTGAGCAAGGTTCTGAAGCGGAAGAGAGAGAGA 180
 DB 141 TTGGCCGCTAGTAGACAGATGCTGAGCAAGGTTCTGAAGCGGAAGAGAGAGAGA 200
 QY 181 GGAGAGAGAACCTCTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 DB 201 GGAGAGAGAACCTCTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 260
 QY 241 ACAGGACACCCCGCGCGCGGTTCTGAGTCCCTTTTACCTCTGAGTCTGAGTCTGCA 300
 DB 261 ACAGGACACCCCGCGCGCGGTTCTGAGTCCCTTTTACCTCTGAGTCTGAGTCTGCA 320
 QY 301 CCACAGCTGAGCAGAGTGAAGCGGACCTGCGGACCTGCTGCTGCTGCTGCTGCTG 360
 DB 321 CCACAGCTGAGCAGAGTGAAGCGGACCTGCGGACCTGCTGCTGCTGCTGCTGCTG 380
 QY 361 GCGGCGATCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
 DB 381 GCGGCGATCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 440
 QY 421 TGACGCCCGCGAGTGGCTGACACTTACTGCAAGCTCGAGCGCTGCTGCTGCTGCTG 480
 DB 441 TGACGCCCGCGAGTGGCTGACACTTACTGCAAGCTCGAGCGCTGCTGCTGCTGCTG 500
 QY 481 CATGGCCAGCTCTGAGAGCCTCAGCAGCATTGAGGCGCTGAGTCAAGCTCCCAACC 540
 DB 501 CATGGCCAGCTCTGAGAGCCTCAGCAGCATTGAGGCGCTGAGTCAAGCTCCCAACC 560
 QY 541 CTTGGCAGACAGGCGCCACAGCGCTGAGCATGGGGGAGCAGCGCCAGCTGGGGTGC 600
 DB 561 CTTGGCAGACAGGCGCCACAGCGCTGAGCATGGGGGAGCAGCGCCAGCTGGGGTGC 620
 QY 601 CTTGGACCTGCTGGGCGCCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 621 CTTGGACCTGCTGGGCGCCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680
 QY 661 TGAGGATATTGACACTCTATGATGACAACTTTGGGACACAGCTTGAAGGCT 720
 DB 681 TGAGGATATTGACACTCTATGATGACAACTTTGGGACACAGCTTGAAGGCT 740
 QY 721 CAACACAGGCGCTGAGATGGGCGCGGAGAGGAAAGCTCGGAGCTGACGAGCGCA 780
 DB 741 CAACACAGGCGCTGAGATGGGCGCGGAGAGGAAAGCTCGGAGCTGACGAGCGCA 800
 QY 781 ATTGAGTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840

DB 801 ATTGAGTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 860
 QY 841 AGGCGCTGAGCCCTCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 DB 861 AGGCGCTGAGCCCTCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 920
 QY 901 GACCACTGCTCTGAGAAAGACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 DB 921 GACCACTGCTCTGAGAAAGACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 980
 QY 961 CACTTTGAGAGACAGATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 DB 981 CACTTTGAGAGACAGATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1040
 QY 1021 CTGCGAGGGGAGCCTGGAATTAACCCCTAGTGAATGATGATGATGATGATGATGATGAT 1080
 DB 1041 CTGCGAGGGGAGCCTGGAATTAACCCCTAGTGAATGATGATGATGATGATGATGATGAT 1100
 QY 1081 GAATTCCTGCGCCCTGGGCTATGACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 DB 1101 GAATTCCTGCGCCCTGGGCTATGACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1160
 QY 1141 CCAATCAGATTTTCAATTAAGCCAGTCTGGGAAATCTCAAA 1185
 DB 1161 CCAATCAGATTTTCAATTAAGCCAGTCTGGGAAATCTCAAA 1205

RESULT 4

AAH99491/C
 ID AAH99491 standard; cDNA: 1181 BP.

AAH99491;

16-OCT-2001 (first entry)

Human protein encoding cDNA sequence SEQ ID NO:326.

Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 anti-inflammatory; antihypertensive; antidiabetic; antidiabetic;
 anti-HIV; fungicide; antimitogen; cardiovascular; antianemic; anemia;
 antiangiogenic; haemostatic; vulvar; antileuk; osteoporosis; eczema;
 dermatological; antiallergic; antidiabetic; antidiabetic; cytostatic;
 neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 genetic disease; haematopoietic disorder; platelet disorder; asthma;
 thrombocytopenia; osteoporosis; severe combined immunodeficiency;
 allergic rhinitis; diabetes; multiple sclerosis; depression;
 Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 neurological disorder; ss.

Homo sapiens.

WO200153455-A2.

26-JUL-2001.

22-DEC-2000; 2000WO-US35017.

23-DEC-1999; 99US-0471275.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

(HSE-) HSE INC.

Tang YT, Liu C, Drmanac RT;

WPI: 2001-457603/49.

P-PSDB; AAM25550.

Isolated human polynucleotides encoding polypeptides, useful for the

PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX Claim 1; Page 441, 1217pp; English.
XX
CC AAH99166 to AAH99904 encode the human proteins given in AAH25225 to
CC AAH25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antineoplastic; antidiabetic;
CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antineoplastic; antiaggregant; haemostatic; vulnery;
CC antitumor; osteoprotic; dermatological; antiallergic; antineoplastic;
CC antidiabetic; cytosolic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SQ Sequence 1181 BP; 226 A; 373 C; 347 G; 235 T; 0 other;

Query Match 97.9%; Score 1167; DB 22; Length 1181;
Best Local Similarity 100.0%; Pred. No. 2.8e-287;
Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 CGCGACCCAGGCGCGGTTTCCGAGAGACGCCAACAGATGTCGCCGCCGCTTTC 85
DB 1169 CGCGACCCAGGCGCGGTTTCCGAGAGACGCCAACAGATGTCGCCGCCGCTTTC 1110
QY 86 CTGATTTGGTTGGGTTGGCTACCTCTTGTCTGATTTGGCGGCTAGTGAAGATGCTG 145
DB 1109 CTGATTTGGTTGGGTTGGCTACCTCTTGTCTGATTTGGCGGCTAGTGAAGATGCTG 1050
QY 146 AGCAAGGCTCTGAACCGGAAACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 205
DB 1049 AGCAAGGCTCTGAACCGGAAACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 990
QY 206 TCCCTGGTGGTATGATCTGGGACGACGAGCGGTGGACAGAGACCCCGGCGCTGGCT 265
DB 989 TCCCTGGTGGTATGATCTGGGACGACGAGCGGTGGACAGAGACCCCGGCGCTGGCT 930
QY 266 AGCTCCCTCTTTGACCTCTCACTGCTCAAGCTCCACACAGCTCTCAGAGAGTGAAGCG 325
DB 929 AGCTCCCTCTTTGACCTCTCACTGCTCAAGCTCCACACAGCTCTCAGAGAGTGAAGCG 870
QY 326 GACCTGGGACACCTGCTGGTGGTGAACACTGGCGGCGATCCAGCGCTCCATGGCA 385
DB 869 GACCTGGGACACCTGCTGGTGGTGAACACTGGCGGCGATCCAGCGCTCCATGGCA 810
QY 386 CCGCGGCGTCCCTCCACACTGCTGCTGACCCAGCTGAGAGCCGCTCTGAGAGAC 445
DB 809 CCGCGGCGTCCCTCCACACTGCTGCTGACCCAGCTGAGAGCCGCTCTGAGAGAC 750
QY 446 TTACTGGCAAGCTGGAGCGCTGCTTTCAAGCTCTCAATGGCAGCTCTCTGAGAGACCTC 505
DB 749 TTACTGGCAAGCTGGAGCGCTGCTTTCAAGCTCTCAATGGCAGCTCTCTGAGAGACCTC 690
QY 506 AGCCACATTTGAGGGCTGAGTCAGGCTCCCAACCTTGGGAGAGAGAGGGGACACAGGC 565
DB 689 AGCCACATTTGAGGGCTGAGTCAGGCTCCCAACCTTGGGAGAGAGAGGGGACACAGGC 630
QY 566 CGTAGCATTCGGGGAGAGAGCGCCAGCTGGGTGGCTTGGAGACTCTGGGGCCACAGCACT 625
DB 629 CGTAGCATTCGGGGAGAGAGCGCCAGCTGGGTGGCTTGGAGACTCTGGGGCCACAGCACT 570
QY 626 GGCTGTCTACTGGAGAGATGGGCTTGAAGGCTGTTGAGAGATATTGACACCTCTATGTAT 685

DB 569 GGCTGTCTACTGGAGAGATGGGCTTGAAGGCTGTTGAGAGATATTGACACCTCTATGTAT 510
QY 686 GACATTAATCTTTGGGACACCACTCTGAGGGGCTTCAACACAGGCCCTTGAAGATGGCCG 745
DB 509 GACATTAATCTTTGGGACACCACTCTGAGGGGCTTCAACACAGGCCCTTGAAGATGGCCG 450
QY 746 GGCAGAGAGAGAGCTCCGAGCTGAGCGAGCGCCGAATTTGATCTCATGATGATGCTG 805
DB 449 GGCAGAGAGAGAGCTCCGAGCTGAGCGAGCGCCGAATTTGATCTCATGATGATGCTG 390
QY 806 GTGGGACACAGGACCTGAGCGAGCCCGGGGCGCAGGCGCTGAGCCCTCTGCTGGA 865
DB 389 GTGGGACACAGGACCTGAGCGAGCCCGGGGCGCAGGCGCTGAGCCCTCTGCTGGA 330
QY 866 TGGTGTCTGATGATGAGCTGAGCCCTGCTGGCTGAGACCACTGCTGAGAGAGAGACA 925
DB 329 TGGTGTCTGATGATGAGCTGAGCCCTGCTGGCTGAGACCACTGCTGAGAGAGAGACA 270
QY 926 GCTGGCTTCCCTAGTACAGAGAGAGAGGCTTGGGCTTGGAGAGAGAGATCTAGTC 985
DB 269 GCTGGCTTCCCTAGTACAGAGAGAGAGGCTTGGGCTTGGAGAGAGAGATCTAGTC 210
QY 986 CTGGGCACTTCAATCTGCTCTCTGCTCAGGGCTGGAGAGGGGAGGCTGGAATTACC 1045
DB 209 CTGGGCACTTCAATCTGCTCTCTGCTCAGGGCTGGAGAGGGGAGGCTGGAATTACC 150
QY 1046 CCTGTGTGATGATGAGAGAGAGGCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 1105
DB 149 CCTGTGTGATGATGAGAGAGAGGCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 90
QY 1106 CTTGGGCTGCTCTCTCTGATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1165
DB 89 CTTGGGCTGCTCTCTCTGATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 30
QY 1166 CAGTCTGGGAAATCTCAAAAAA 1192
DB 29 CAGTCTGGGAAATCTCAAAAAA 3

RESULT 5
AAC76627/c
ID AAC76627 standard; cDNA; 1222 BP.
XX
AC AAC76627;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORF2182 polynucleotide sequence SEQ ID NO:4363.
XX
XX
KW Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
KW anticonvulsant; osteopathic; antiparkinsonian; nootropic; neuroprotective;
KW antineoplastic; antiallergic; antidiabetic; immunosuppressive; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineoplastic; antihypertensive;
KW antineoplastic; antidiabetic; antineoplastic; antineoplastic; antineoplastic;
KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antineoplastic disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
PN WO200058473-A2.
XX
XX 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-0508621.
XX

PR 31-MAR-1999; 9905-0127607.
 PR 02-APR-1999; 9905-0127636.
 PR 05-APR-1999; 9905-0127728.
 PR 30-MAR-2000; 200005-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB42418.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 3559; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferic; antiparkinsonian; nootropic; neuroprotective;
 CC osteoprotic; anticonvulsant; antirheptic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertoid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 1222 BP; 229 A; 370 C; 350 G; 270 T; 3 other:
 Query Match 96.8%; Score 1154.4; DB 21; Length 1222;
 Best Local Similarity 99.7%; Pred. No. 4.5e-284;
 Matches 1178; Conservative 0; Mismatches 1; Indels 3; Gaps 2:
 QY 14 AGAAGTGGTCTGCGGACCAAGGCGGGTTTCGGAGAGACCAACAGCATGTGCG 73
 DB 1217 AGAAGTGGTCTGCGGACCAAGGCGGGTTTCGGAGAGACCAACAGCATGTGCG 1158
 QY 74 CGCGCGCTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATG 133
 DB 1157 CGCGCGCTTCCGATGATGATGATGATGATGATGATGATGATGATGATG 1098
 QY 134 AGCAAGATGCTGAGCAAGGCTCTGAAGCGGAAACGAGAGAGAGAGAGAGAACT 193
 DB 1097 AGCAAGATGCTGAGCAAGGCTCTGAAGCGGAAACGAGAGAGAGAGAGAACT 1038
 QY 194 CTGGCAGTCACTCCTGCTGATGATGATGATGATGATGATGATGATGATGATG 253
 DB 1037 CTGGCAGTCACTCCTGCTGATGATGATGATGATGATGATGATGATGATGATG 978
 QY 254 GCGGCGGCTCTAGTCCCTCTTTGACCTCTCAAGGCTCAAGGCTCC-ACCAAGCTTGA 312
 DB 977 GCGGCGGCTCTAGTCCCTCTTTGACCTCTCAAGGCTCAAGGCTCC-ACCAAGCTTGA 918
 QY 313 GCAGAGTACGCGGACCTGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 372
 DB 917 GCAAGTACGCGGACCTGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 858
 QY 373 --GGGCTCATGAGACCGCGGCTGCCCTGACCTGTGCTTACCTGACCTGACGCCCC 430
 DB 857 NNGCGCTCATGAGACCGCGGCTGCCCTGACCTGTGCTTACCTGACCTGACGCCCC 798

QY 431 AGTGTGCTGACAACTTACTGTGCAAGCTCGGACGCTGCCCTTTCAGCTTCATGCGCAGC 490
 DB 797 AGTGTGCTGACAACTTACTGTGCAAGCTCGGACGCTGCCCTTTCAGCTTCATGCGCAGC 738
 QY 491 CTCTGAGAGACCTTACGACCATTTGAGGCTCTGAGTACGCTGCCCAACCTTTGGCAGAC 550
 DB 737 CTCTGAGAGACCTTACGACCATTTGAGGCTCTGAGTACGCTGCCCAACCTTTGGCAGAC 678
 QY 551 GAGGAGCCACAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 610
 DB 677 GAGGAGCCACAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 618
 QY 611 CTGGGCGGACGACCTGCTGCTTACTGAGCAGATGAGGCTTTGAGGCTTTGAGGATATT 670
 DB 617 CTGGGCGGACGACCTGCTGCTTACTGAGCAGATGAGGCTTTGAGGCTTTGAGGATATT 558
 QY 671 GACACCTTATGATGACAAATGAACTTTGGGACACCAAGCTCTGAGGCTTCAACCAAGC 730
 DB 557 GACACCTTATGATGACAAATGAACTTTGGGACACCAAGCTCTGAGGCTTCAACCAAGC 498
 QY 731 CCTGAGATGAGGCGGCGGACAGAGAGCTCCGAGCTGAGAGAGCGGCAATTTGACTAC 790
 DB 497 CCTGAGATGAGGCGGCGGACAGAGAGCTCCGAGCTGAGAGAGCGGCAATTTGACTAC 438
 QY 791 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 850
 DB 437 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 378
 QY 851 GCCGCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 910
 DB 377 GCCGCGGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 318
 QY 911 CCTGAAAGACAGAGCTGCTTCCCTGATGACAGAGAGAGAGAGAGAGAGAGAGAG 970
 DB 317 CCTGAAAGACAGAGCTGCTTCCCTGATGACAGAGAGAGAGAGAGAGAGAGAGAG 258
 QY 971 AGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1030
 DB 257 AGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 198
 QY 1031 GAGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1090
 DB 197 GAGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 138
 QY 1091 GCCCTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1150
 DB 137 GCCCTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 78
 QY 1151 TTTTCAATTAAGCAAGCTGCTGAGGAAATCTCAAAAAA 1192
 DB 77 TTTTCAATTAAGCAAGCTGCTGAGGAAATCTCAAAAAA 36
 RESULT 6
 AAX97688
 ID AAX97688 standard; DNA; 1182 BP.
 AAX97688;
 DT 13-SEP-1999 (first entry)
 DE Extended human secreted protein coding sequence, SEQ ID NO. 253.
 XX
 KW Secreted protein; human; cytokine; cellular proliferation; cell movement;
 KW cellular differentiation; immune system regulator; anti-inflammatory;
 KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
 KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
 KW genetic disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO9931236-A2.

This sequence represents an extended human secreted protein coding sequence of the invention. The secreted proteins can be used in treating or controlling a variety of human conditions. The secreted proteins may act as cytokines or may affect cellular proliferation or differentiation or may act as immune system regulators, haematopoietic regulators, tissue growth regulators, regulators of reproductive hormones or cell movement or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or tumour inhibition activity. The DNAs can be used in forensic procedures to identify individuals or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal expression of the genes corresponding to the extended cDNAs. They are also useful for constructing a high resolution map of the human chromosomes. They can also be used for gene therapy to control or treat genetic diseases.

SQ Sequence 1182 BP; 239 A; 348 C; 357 G; 227 T; 11 other.

Query Match	96.6%	Score 1151.2	DB 20	Length 1182
Best Local Similarity	98.6%	Pred. NO. 2.9e-283		
Matches 1102, Conservative	8	Mismatches 8	Indels 1	Gaps 1

OY	1	AGAACTGCGCTTGGCAACCCAGGCGGGGTTTCGGAGAGACGCCAACCAACGATGGCTGC	73
Db	2	AGAACTGCGCTTGGCAACCCAGGCGGGGTTTCGGAGAGACGCCAACCAACGATGGCTGC	61
OY	74	CGCCGCGTTTCCGTGATGGTGTGGGTGGCTACCTCTCGTTTGATTGGCCGCTAGTG	133
Db	62	CGCCGCGGTTTCCGTGATGGTGTGGGTGGCTACCTCTCGTTTGATTGGCCGCTAGTG	121
OY	134	AGCAAGTGTGTGCAAGGGTCTGAAGCGGAAACGGGAGAGAGAGAGAGAGAGAACT	193
Db	122	AGCAAGTGTGTGCAAGGGTCTGAAGCGGAAACGGGAGAGAGAGAGAGAGAAAGAACT	181
OY	194	CTGGCAGTGCATCTCTGTTGGCTAGTCTCTGGCCACGAGAGGGTGGCACAGGCAACCCCG	253
Db	182	CTGGCAGTGCATCTCTGTTGGCTAGTCTCTGGCCACGAGAGGGTGGCACAGGCAACCCCG	241
OY	254	GCCGTGGCCTTAGCTCCCTCTTTGACCTCTCAGTGTCAAGCTCACACAGCTCTCAG	313
Db	242	GCCGTGGCCTTAGCTCCCTCTTTGACCTCTCAGTGTCAAGCTCACACAGCTCTCAG	301
OY	314	CAGAGTGAAGCCGACCTTGGGCACTGGGAGCTGGTGGTGAACACTCTGGCGGCAATCAG	373
Db	302	VRAAGTGAAGCCGACCTTGGGCACTGGTGGTGGTGAACACTCTGGCGGCAATCAG	361
OY	374	GCGTCATGGCAACCGCGGCTGCCTTGCACCTGTGCTAGCCCACTGACGACCCCAAGT	433
Db	362	GCGTCATGGCAACCGCGGCTGCCTTGCACCTGTGCTAGCCCACTGACGACCCCAANT	421
OY	434	GTTGGTGAACAATTATGGCAGATCGGACGCTGGCTTGAAGCTTCATATGGCCAGATC	493
Db	422	GTTGGTGAACAATTATGGCAGATCGGACGCTGGCTTGAAGCTTCATATGGCCAGATC	481

OY	494	TTGAGGAGACCTCAGCCCAATGATGAGGCGCTGAGTACAGGCTCCCAACCTCTGGCAACGAG	553
Db	482	CTGGARGACCTCAGCCCAATGATGAGGCGCTGAGTACAGGCTCCCAACCTCTGGCAACGAG	541
OY	554	GGGCGACAGGAGCCGTAGCATTGAGGGGAGCAGCGCCAGCTGAGGAGCGCTTGGACCTGAGC	613
Db	542	GGGCGACAGGAGCCGTAGCATTGAGGGGAGCAGCGCCAGCTGAGGAGCGCTTGGACCTGAGC	601
OY	614	GGCCGAGCCACTGAGCTGTCTACTGAGCAGATGGGCTTGAAGGGCGCTTTGAGATATTGAC	673
Db	602	GGCCGAGCCACTGAGCTGTCTACTGAGCAGATGGGCTTGAAGGGCGCTTTGAGATATTGAC	661
OY	674	ACCTGTATGTATGACATGAACTTTGGGACACCAGCTTGTGAGGGCGCTTAACACGAGCGCT	733
Db	662	ACCTGTATGTATGACATGAACTTTGGGACACCAGCTTGTGAGGGCGCTTAACACGAGCGCT	721
OY	734	GAGGATGGGGCGGGCAAGGAGGAACTCCGAGCTGGAGCAGAGGCCGGAATTGGACTACTCTC	793
Db	722	GAGGATGGGGCGGGCAAGGAGGAACTCCGAGCTGGAGCAGAGGCCGGAATTGGACTACTCTC	781
OY	794	ATGATGTGCTGAGTGGGACACACAGGACCTGGAGCAGCAGCCGGGGCCAGAGGGCGTGAGGC	853
Db	782	ATGATGTGCTGAGTGGGACACACAGGACCTGGAGCAGCAGCCGGGGCCAGAGGGCGTGAGGC	841
OY	854	CTCGTCTGGAATGGTGTCTGATCTGAACTGAGCCTGCTGAGTGGACCAACTGTCTCT	913
Db	842	CTCGTCTGGAATGGTGTCTGATCTGAACTGAGCCTGCTGAGTGGACCAACTGTCTCTCT	901
OY	914	CGAAAAGACACAGCTGGCTTCCCTTAGTACAGAGAACAGGGCTTGGGCCACTTTGGACAGA	973
Db	902	CGAAAAGACACAGCTGGCTTCCCTTAGTACAGAGAACAGGGCTTGGGCCACTTTGGACAGA	961
OY	974	CAGAACTAGTCTGGGGACCTTCACATCCGCGCTGTCAGAGGCTGGGACAGGGGAG	1033
Db	962	CAGAACTAGTCTGGGGACCTTCACATCCGCGCTGTCAGAGGCTGGGACAGGGGAG	1021
OY	1034	CCTGGAATTACCCCTAGTGTGATGAGATGACAGGGTCTGTGGGGAGCTGAATTTCCCTGGCC	1093
Db	1022	CCTGGAATTACCCCTAGTGTGATGAGATGACAGGGTCTGTGGGGAGCTGAATTTCCCTGGCC	1081
OY	1094	CTGGGGCTATAGCTTGGGCTGTCTTCTCTGATACGGGAGAGACCCCAATCAGATTTT	1155
Db	1082	CTGGGGCTATAGCTTGGGCTGTCTTCTCTGATACGGGAGAGAG-CCCAATCAGATTTT	1144
OY	1154	TCAAATTTAAAGCAGCTCTGGGAAATCCAAAAAATAA	1192
Db	1141	TCAAATTTAAAGCAGCTCTGGGAAATCCAAAAAATAA	1179

RESULT 7	
AAZ19894	
ID	AAZ19894 standard; cDNA; 1078 BP.
XX	
AC	AAZ19894;
XX	
DT	06-DEC-1999 (first entry)
XX	
DE	Human adult blood secreted protein g21_1 cDNA.
XX	
KM	Secreted protein; g21_1; human; therapy; diagnosis; vaccine;
KW	blood; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	
FT	Location/Qualifiers
FT	10..720
FT	/*tag= a
FT	/note= "this region is specifically claimed in
FT	Claim 12(b)"
FT	10..246
FT	/*tag= b
FT	247..717
FT	mat_peptide
FT	/*tag= c

```

FT      /note= "this region (plus the stop codon) is
FT      sig_peptide 10..102 specifically claimed in Claim 12(c) "
FT      /tag= d
FT      /note= "alternative signal sequence"
FT      mat_peptide 103..717
FT      /tag= e
FT      /note= "alternative mature protein"
XX      WO947555-A1.
XX      23-SEP-1999.
XX      18-MAR-1999; 99WO-US05939.
XX      20-MAR-1998; 98US-0078803.
XX      17-MAR-1999; 99US-0078803.
XX      (GENY ) GENETICS INST INC.
XX      Jacobs K, McCoy JM, LaValle ER, Collins-Racie LA, Evans C;
XX      Merberg D, Treacy M, Agostino MJ, Steinger RJ;
XX      P-PSDB; AAY31829.
XX      WPI: 1999-562059/47.
XX      P-PSDB; AAY31829.
XX      New polynucleotides derived from murine foetal cell cDNA libraries,
XX      potentially used as, e.g. vaccines -
XX      Claim 12(a); Page 93; 107pp; English.
XX      This is the nucleotide sequence of an isolated cDNA clone that
XX      codes for a novel human secreted protein, g21_1 (see AAY31829).
XX      The clone was isolated from a human adult blood (peripheral blood
XX      mononuclear cells treated with phytohemagglutinin, phorbol
XX      myristate acetate and mixed lymphocyte reaction) cDNA library
XX      using human cDNAs that had themselves been identified in a human
XX      adult blood cDNA library using methods selective for cDNAs encoding
XX      secreted proteins. The clone is deposited in composite deposit
XX      ATCC 99700. The invention provides new human secreted proteins
XX      (see AAY31828-38) and polynucleotides (see AAY19893-901) isolated from
XX      foetal cell, adult blood, adult brain and foetal kidney cDNA
XX      libraries. They are predicted to have biological activities which
XX      would make them suitable for treating, preventing or ameliorating
XX      medical conditions in humans and animals, although no supporting
XX      data are given. Suggested activities include nutritional,
XX      cytokine, tissue growth, cell proliferation and differentiation,
XX      immunostimulant (e.g. as vaccine), immunosuppressive,
XX      haematopoiesis regulating, activin or inhibin, chemotactic or
XX      chemokinetic, haemostatic or thrombolytic, receptor/ligand activity,
XX      antiinflammatory, cadherin or tumour invasion suppressor, and
XX      tumour inhibition activities. The polynucleotides are also stated
XX      to be useful for gene therapy.
XX      Sequence 1078 BP: 234 A; 311 C; 332 G; 199 T; 2 other:
SQ
Query Match 89.0%; Score 1061.2; DB 20; Length 1078;
Best Local Similarity 99.8%; Pred. No. 2.4e-260;
Matches 1060; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      |||||
181 CAGCAGATGAGCCGACCTGGGACCTGGTGTGCTGTAACACTCTCGGGGCATC 240
QY      CAGGCGTCATAGGCACCCGCGGCTGCTGGCACCTGTGCGTACGACCTGACGCCCC 430
Db      |||||
241 CAGGCGTCATAGGCACCCGCGGCTGCTGGCACCTGTGCGTACGACCTGACGCCCC 300
QY      AGTGTGCTGACAACTACTGAGCAAGTGGACGCTGCTTTCAGCTTCATGGCCAGC 490
Db      |||||
301 AGTGTGCTGACAACTACTGAGCAAGTGGACGCTGCTTTCAGCTTCATGGCCAGC 360
QY      CTCTGAGAGACCTTAGCCACATTGAGGCTGAGTCAAGCTCCCAACCTTGGCAGAC 550
Db      |||||
361 CTCTGAGAGACCTTAGCCACATTGAGGCTGAGTCAAGCTCCCAACCTTGGCAGAC 420
QY      GAGGGGCCACACAGGCGCTAGCATCGGGGAGACAGCGCCAGCCTGGTGGACCTG 610
Db      |||||
421 GAGGGGCCACACAGGCGCTAGCATCGGGGAGACAGCGCCAGCCTGGTGGACCTG 480
QY      CTGGGCCACGCACTGGCTGTCTACTGAGCAGATGGGCTTGAGGGCTTTGAGGATAT 670
Db      |||||
481 CTGGGCCACGCACTGGCTGTCTACTGAGCAGATGGGCTTGAGGGCTTTGAGGATAT 540
QY      GACACCTCTATGTATGACAAATGAACTTTGGGACACGCTCTGAGGGCTTCAAAACAGG 600
Db      |||||
601 GACACCTCTATGTATGACAAATGAACTTTGGGACACGCTCTGAGGGCTTCAAAACAGG 660
QY      CCTGAGATGGGCGCGGACAGAGAGACCTCGAGCTGAGCAGAGCCGAATTTGAGATAC 790
Db      |||||
791 CTGATGATGTGCTGTGTTGGGACACAGGCACTGAGAGCAGCGCGGCGGCGCTGA 850
QY      CTGATGATGTGCTGTGTTGGGACACAGGCACTGAGAGCAGCGCGGCGGCGCTGA 850
Db      |||||
851 GCGGCGGCTGAGTATGCTGTGCTGTAAGTGAAGCTGCTGGGTGAGCAACTGT 910
QY      GCGGCGGCTGAGTATGCTGTGCTGTAAGTGAAGCTGCTGGGTGAGCAACTGT 780
Db      |||||
911 CCTGGAAGACACAGCTGCGCTTCCCTAGTACAGAGACAGGCGCTTGGGCACTTTGGAG 970
QY      CCTGGAAGACACAGCTGCGCTTCCCTAGTACAGAGACAGGCGCTTGGGCACTTTGGAG 970
Db      |||||
971 AGACGAATCTAGTCTGTTGGGCACTTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1030
QY      AGACGAATCTAGTCTGTTGGGCACTTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db      |||||
1031 GAGCCTGGAATTAACCCCTAGTATGATGATGATGATGATGATGATGATGATGATGATG 1090
QY      GAGCCTGGAATTAACCCCTAGTATGATGATGATGATGATGATGATGATGATGATGATG 960
Db      |||||
961 GCGGCGGCTGATATCTGTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1150
QY      GCGGCGGCTGATATCTGTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db      |||||
1151 TTTTCAATTTAAGCAGTCTGAGGAATTCATAAAAAA 1192
QY      TTTTCAATTTAAGCAGTCTGAGGAATTCATAAAAAA 1062
Db      |||||
1021 TTTTCAATTTAAGCAGTCTGAGGAATTCATAAAAAA 1062

```

RESULT 8
AAS78542/c
ID AAS78542 standard; cDNA, 1216 BP.
XX AAS78542;
AC AAS78542;
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #14346.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.
XX NC200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001: 2001MO-0508631.
XX
XX 31-MAR-2000: 2000US-0540217.
PR 23-AUG-2000: 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT:
XX
DR WPI: 2001-639362/73.
XX
XX P-PSDB: ABG14355.
PT
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 1: SEQ ID NO 14346; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAs64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1216 BP; 239 A; 379 C; 345 G; 253 T; 0 other:
Query Match 74.6%; Score 889.4; DB 23; Length 1216;
Best Local Similarity 93.0%; Pred. No. 1.4e-216;
Matches 1111; Conservative 0; Mismatches 56; Indels 28; Gaps 16;
QY 26 CGGACCCAGGCGGGGTTTCGGAGAGACCAACAGAGTGGCCGCCCTTTC 85
DB 1204 CGGACCCAGGCGGGGTTTCGGAGAGACCAACAGAGTGGCCGCCCTTTC 1145
QY 86 CTGATTGATTGGGCTGCTACCTCTTCCTTCTGATTGGCCGCTAGTGAAGATGCTG 145
DB 1144 CTGATTGATTGGGCTGCTACCTCTTCCTTCTGATTGGCCGCTAGTGAAGATGCTG 1085
QY 146 AGCAAGGCTCTGAAGCGGAAGCGAGAGAGAGAGAGAGAGTCTGGCAGTGCAC 205
DB 1084 AGCAAGGCTCTGAAGCGGAAGCGAGAGAGAGAGAGAGAGTCTGGCAGTGCAC 1025
QY 206 TCCGTGGCTAGATCTCTGGCCAGCAGAGGGTGGCAGAGCACCCTGGGCGGTGCTCT 265
DB 1024 TCCGTGGCTAGATCTCTGGCCAGCAGAGGGTGGCAGAGCACCCTGGGCGGTGCTCT 965
QY 266 -AGCTCCCTCTTGAAGCTCTCAAGTCTCAAGCTCCACAC -AGCTCTCAAGAGTGA - 321
DB 964 AAGCTCCCTCTTGAAGCTCTCAAGTCTCAAGCTCCACACAGCTCTGAGAGAGTGAAG 905

QY 322 GCCGAGCTGCGGACCTGCTGTG-TGCTGAACACTCTGCGGCATCCAGAGC--TC 378
DB 904 CCGAGACCTTGCGGACCTGCTGTGCTGAAGCTCTGCGGCATCCAGAGTCTGCG 845
QY 379 CATGGACCCCGGGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 438
DB 844 TTTGAACCCCGGGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 785
QY 439 TGAC-AAGTTACTGCGAAGCTGCGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 497
DB 784 TGACTAATCTAGCTGCGAAGCTGCGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 725
QY 498 AGGACCTGACCAATTTGAGGGCTGAGTCCAGGCTCCCAACCTTGGAGAGAGAGAGG 557
DB 724 AGGACCTGACCAATTTGAGGGCTGAGTCCAGGCTCCCAACCTTGGAGAGAGAGAGG 665
QY 558 CACGAGGC-GTAGCATCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 616
DB 664 CCAGAGCCAGTAGATCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
QY 617 CCAGCCACTGCGTGTCTACTGAGAGATGGGCTTGAAGGCTGTTGAGATATTGACACC 676
DB 604 CCAGCCACTGCGTGTCTACTGAGAGATGGGCTTGAAGGCTGTTGAGATATTGACACC 545
QY 677 TCTATGATGACATGAC-TTTGGGACACAGCTCT-GAGGGCTCAACACAGGCTCTG 734
DB 544 TCTATGATGACATGACATTTTGGGACACAGCTCTGAGGGCTCAACACAGGCTCTG 485
QY 735 AGGATGG---CGGGCAAG 790
DB 484 AAGAGTGGCCCGGGGAG 425
QY 791 CTCATGATGCTGTGTGGGACACAGAGCA--CTGAGAGAGAGAGAGAGAGAGAGAGAGAG 846
DB 424 CTCATGATGCTGTGTGGGACACAGAGCACTGAGAGAGAGAGAGAGAGAGAGAGAGAG 365
QY 847 CTGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 906
DB 364 CTGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 305
QY 907 CTTGCTCTGCAAAAAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 965
DB 304 CTTGCTCTGCAAAAAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245
QY 966 TGGAGAGACAGATCTAG--TCCGTGGCACTTCACATCCGCTCTGCTGCTGCTGCTGCT 1017
DB 244 TGGAGAGACAGATCTAGTTCCTGGGCACTTCACATCCGCTCTGCTGCTGCTGCTGCT 185
QY 1018 GGGCTGGCAGGGGAGCTGGAATTAACCCCTAGTATGATGAATGACAGGCTCTGCTG 1077
DB 184 CGGCAAGGGGGAGGCTTGAATTAACCCCTAGTATGATGAATGACAGGCTCTGCTG 125
QY 1078 ACTGAATTCCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1137
DB 124 ACTGAATTCCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 65
QY 1138 ACCCAATGAGATTTTCAATTAAGCCAGTCTGGAATTAAGTCAAAAAA 1192
DB 64 ACCCAATGAGATTTTCAATTAAGCCAGTCTGGAATTAAGTCAAAAAA 10
RESULT 9
ABK81819
ID ABK81819 standard; DNA: 5746 BP.
XX
XX ABK81819;
AC
AC 13-AUG-2002 (first entry)
XX
XX DNA representing lung specific gene #5.
DE Lung specific gene; gene therapy; vaccine; lung cancer;
XX cancer staging; cancer monitoring; cancer diagnosis;
KW

Query Match 59.4%; Score 707.8; DB 20; Length 711;
 Best Local Similarity 99.7%; Pred. No. 2,2e-170;
 Matches 709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 140 ATGCTAGCAAGAGGTGTGAAGACGGAGAGAGAGAGAGAGAGAACTCTGGCA 199
 |||||||
 DB 1 ATCTTAGCAAGAGGTGTGAAGACGGAGAGAGAGAGAGAGAGAACTCTGGCA 60
 QY 200 GTGCACTCTGTGTAGTATCTTGCCACAGCAGCGGTGGACACAGCCCGGCGTG 259
 |||||||
 DB 61 GTGCACTCTGTGTAGTATCTTGCCACAGCAGCGGTGGACACAGCCCGGCGTG 120
 QY 260 GCCTTAGCTCCCTCTTTGACTCTGAGTGTCAAGCTCCACAGCAGCTGACAGAGT 319
 |||||||
 DB 121 GCCTTAGCTCCCTCTTTGACTCTGAGTGTCAAGCTCCACAGCAGCTGACAGAGT 180
 QY 320 GAGCCGAGCTGGGGAGCCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 379
 |||||||
 DB 181 GAGCCGAGCTGGGGAGCCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
 QY 380 ATGGACCCGCGGCTCCCTGACACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 439
 |||||||
 DB 241 ATGGACCCGCGGCTCCCTGACACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
 QY 440 GACAACTTACTGGCAAGCTGGACGCTGCCCTTTGAGCTCCATGGCCAGCCTCTGAG 499
 |||||||
 DB 301 GACAACTTACTGGCAAGCTGGACGCTGCCCTTTGAGCTCCATGGCCAGCCTCTGAG 360
 QY 500 GACCTGACCCACATTTAGGAGCCTGAGTCAAGGCTCCCAACCTTTGGCAGAGGGGCA 559
 |||||||
 DB 361 GACCTGACCCACATTTAGGAGCCTGAGTCAAGGCTCCCAACCTTTGGCAGAGGGGCA 420
 QY 560 CCAGGCGGTAGCATCGGGGAGAGAGCGCCAGCCTGGTGTGTGTGTGTGTGTGTGTGT 619
 |||||||
 DB 421 CCAGGCGGTAGCATCGGGGAGAGAGCGCCAGCCTGGTGTGTGTGTGTGTGTGTGTGT 480
 QY 620 GCCACGT 679
 |||||||
 DB 481 GCCACGT 540
 QY 680 ATGTATGACAAATGAACTTTGGGACACAGCCTGTGAGGGGCTCAAAACAGGCGCTGAGAT 739
 |||||||
 DB 541 ATGTATGACAAATGAACTTTGGGACACAGCCTGTGAGGGGCTCAAAACAGGCGCTGAGAT 600
 QY 740 GGGCGGCGCAAGAGAGAGCTCCGAGCTGGAGCTGGAGAGCGCAATTTGACTTACCTCATGAT 799
 |||||||
 DB 601 GGGCGGCGCAAGAGAGAGCTCCGAGCTGGAGCTGGAGAGCGCAATTTGACTTACCTCATGAT 660
 QY 800 GTGCTGT 850
 |||||||
 DB 661 GTGCTGT 711

RESULT 11
 AAL49936 standard; cDNA; 5869 BP.
 AC AAL49936;
 XX 10-DEC-2002 (first entry)
 DE Human molecule for disease detection and treatment coding sequence #10.
 XX Human; molecule for disease detection and treatment; MDR; gene therapy;
 KW cytostatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;
 KM antiinflammatory; antiashmatic; cerebroprotective; nootropic;
 KM neuroprotective; antiparkinsonian; cardiac; antidiabetic; gene; ss.
 XX Homo sapiens.
 OS
 XX
 PN MO200270709-A2.
 XX
 PD 12-SEP-2002.

XX 08-FEB-2002; 2002MO-US03709.
 PF
 XX
 PR 09-FEB-2001; 2001US-268117P.
 PR 15-FEB-2001; 2001US-269618P.
 PR 23-FEB-2001; 2001US-271118P.
 PR 07-MAR-2001; 2001US-274486P.
 PR 09-MAR-2001; 2001US-274436P.
 PR 28-NOV-2001; 2001US-334229P.
 PR 01-FEB-2002; 2002US-353284P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lal PG, Baughn MR, Yao MG, Walla NK, Elliot VS, Xu Y;
 PI Honcheil CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DM;
 PI Hafalia AJA, Ghadi AR, Thangavelu K, Sanjanna MM, Tang YT;
 PI Ramkumar J, Griffith JA, Swannaker A, Azimati Y, Sapperstein SK;
 PI Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;
 XX
 DR MPI; 2002-713453/77.
 DR P-PSDB; AAO19407.
 XX
 PT New human molecules for disease detection and treatment (MDR), useful
 PT for diagnosing, treating and preventing diseases or conditions
 PT associated with the aberrant MDR expression, e.g. cancer, AIDS,
 PT asthma, diabetes, hepatitis -
 XX
 PS Claim 11; Page 166-167; 177pp; English.
 XX
 CC The present invention relates to human proteins and coding sequences of
 CC molecules for disease detection and treatment MDR. The sequences can be
 CC used in the treatment of diseases associated with the decreased
 CC expression or overexpression of MDR, such as cell proliferative (cancer,
 CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,
 CC allergies, Addison's disease, asthma), developmental (dwarfism, renal
 CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,
 CC epilepsy) and cardiovascular (congestive heart failure, myocardial
 CC infarction, angina pectoris) disorders. The present sequence is a
 CC coding sequence of the invention.
 XX
 SQ Sequence 5869 BP; 1191 A; 1682 C; 1995 G; 1001 T; 0 other;

Query Match 56.0%; Score 667.4; DB 24; Length 5869;
 Best Local Similarity 99.9%; Pred. No. 8.5e-160;
 Matches 668; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 138 AGATGTGAGCAAGGCTGTGAAGCGGAACGGAGAGAGAGAGAGAGAACTCTGG 197
 |||||||
 DB 237 AGATGTGAGCAAGGCTGTGAAGCGGAACGGAGAGAGAGAGAGAACTCTGG 296
 QY 198 CAGTCGACCTCTGT 257
 |||||||
 DB 297 CAGTCGACCTCTGT 356
 QY 258 TGGCTCTAGCTCCCTCTTTGACTCTGAGTGTCAAGCTCCACAGCAGCTGACAGAGA 317
 |||||||
 DB 357 TGGCTCTAGCTCCCTCTTTGACTCTGAGTGTCAAGCTCCACAGCAGCTGACAGAGA 416
 QY 318 GTGAGCGGAGACCTCGCGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 377
 |||||||
 DB 417 GTGAGCGGAGACCTCGCGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 476
 QY 378 CCATGGCACCGGCGGCTGT 437
 |||||||
 DB 477 CCATGGCACCGGCGGCTGT 536
 QY 438 CTGACAACTTACTGTGCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 497
 |||||||
 DB 537 CTGACAACTTACTGTGCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 596
 QY 498 AGGACCTGAGCCACATTTGAGGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 557
 |||||||
 DB 597 AGGACCTGAGCCACATTTGAGGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 656

QY	Db
558	517
657	716
618	677
717	776
678	737
777	836
738	797
837	896
QY	Db
798	857
897	905

RESULT 12
AAK91707
ID AAK91707 standard; cDNA; 696 BP

AC AAK91707;

DT 06-NOV-2001 (first entry)

Human cDNA 5'-end sequence, SEQ ID NO: 167.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

PN EP1130094-A2.

05-SEP-2001.

07-JUL-2000; 2000EP-0114089.

PR	08-JUL-1999;	99JP-0194486.
DE	11-JAN-2000	2000TD-0110574

PR 02-MAY-2000; 2000JF-0183/b5-

FA (HELL-) HELIA RES INST.

PI Old I, NISHIKAWA I, 1
Makematau & Suediwee

XX
DB WPT: 2001-534355/EO

030 Performance [for each] Joseph

use in genetic manipulation -

Sequence 696 BP; 120 A; 221 C; 220 G; 132 T; 3 other; CC
The invention relates to primers for synthesising full length cDNA CC
clones. 830 cDNA molecules encoding a human protein have been CC
isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA CC
molecules have been determined. Primers for synthesising the full length CC
cDNA are useful for clarifying the function of the protein encoded by CC
the cDNA. The full length clones were obtained by construction of full CC
length enriched cDNA libraries that were synthesised by the oligo-capping CC
method. The primers enable the production of the full length cDNA easily CC
without any special methods. The present sequence is the nucleotide CC
sequence of the 5'-end of a cDNA provided in the invention. CC
Note: The sequence data for this patent did not form part of the printed CC
specification, but was obtained in CD-ROM format directly from EPO. CC

Sequence 696 BP; 120 A; 221 C; 220 G; 132 T; 3 other;

Query Match	53.5%	Score 637.2;	DB 22;	Length 696;
Best Local Similarity	98.1%;	Pred. No. 2.1e-152;		
Matches 664; Conservative	0;	Mismatches 11;	Indels 2;	Gaps 2

QY	1	GTAGGCGGAGTGTAGAAACTGCGTCTGCGACCCAGGCGGGGTTCGAGAGACGCCAA	60
Db	21	GTAGGCGGAGTGTAGAAACTGCGTCTGCGACCCAGGCGGGGTTCGAGAGACGCCAA	80
QY	61	CAAGCATGCTGCCCGCCGCTTTCCTGATTTGGTGTGGGTGACTCTTTCCTTGA	120
Db	81	CAAGCATGCTGCCCGCCGCGCTTTCCTGATTTGGTGTGGGTGACTCTTTCCTTGA	140
QY	121	TTAGCGCGTAGTAGCAAGATGATGTAGCAAGGGGTCTGAAGCGGAAACGAGAGAGAGA	180
Db	141	TTGGCGCGTAGTAGCAAGATGCTGAGCAAGGGGTCTGAAGCGGAAACGAGAGAGAGA	200
QY	181	GGAGAAAGAACTCTGGCAGTGCACCTCTGTGGCTAGTATGATCTGGCAAGCGCGTGGC	240
QY	201	GGAGAAAGAACTCTGGCAGTGCACCTCTGTGGCTAGTATGATCTGGCAAGCGCGTGGC	260
Db	241	ACAGGACCCCGGGCGGTGGGCTCTAGCTCCCTTTGACCTCTAGTGTCTCAAGCTTCA	300
Db	261	ACAGGACCCCGGGCGGTGGGCTCTAGCTCCCTTTGACCTCTAGTGTCTCAAGCTTCA	320
QY	301	CCACAGCCTGCAGCAGAGTAGAGCCGGACCTTGGCGACCTGTGTGCTGTGTGAACACTCT	360
Db	321	CCACAGCCTGCAGCAGAGTAGAGCCGGACCTTGGCGACCTGTGTGCTGTGTGAACACTCT	380
QY	361	GGGGCCGATTCAGGGGCTCCATGGAGCAACCGCGGCTGCCCTGGCACTGTGTAGCCAC	420
Db	381	GGGGCCGATTCAGGGGCTCCATGGAGCAACCGCGGCTGCCCTGGCACTGTGTAGCCAC	440
QY	421	TGAGGCCCCAGTGTGGTGTGACAACTTATGTGCAAGCTTGGAGCGCTGGCTTTAGGCTC	480
Db	441	TGAGGCCCCAGTGTGGTGTGACAACTTATGTGCAAGCTTGGAGCGCTGGCTTTAGGCTC	500
QY	481	CATGGCCAGCTCTCTGAGAGGACCTTAGCCACATTAGGAGGCTGAGTCAAGGCTCCCAAC	540
Db	501	CATGGCCAGCTCTCTGAGAGGACCTTAGCCACATTAGGAGGCTGAGTCAAGGCTCCCAAC	560
QY	541	CTTGGCAGAGAGAGGGGCCACAGGCGGTATGATCTGGGGAGAGAGGCGCCACTTGGGGTC	600
Db	561	CTTGGCAGAGAGAGGGGCCACAGGCGGTATGATCTGGGGAGAGAGGCGCCACTTGGGGTC	620
QY	601	CTTGGACCTGTGGGGCCAGGCCACTGGCTGTCTACTGAGACGATGGGCTTGAAGGCGCTTT	660
Db	621	CTTGGACCTGTGGGGCCAGGCCACT-GCTTCTACTGAGACAT-GGCTTGAAGNCGTGT	679
QY	661	TGAGATATTTGACACT 677	
Db	679	TGANGATTTTGGCCCTT 695	

RESULT 13
AAK93731
ID AAK93731 standard; cDNA; 696 BP.

AC AAK93731;

DT 06-NOV-2001 (first entry)

DE Human cDNA clone representative sequence, SEQ ID NO: 2191.

KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

05 Homo sapiens.

PN EP1130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.


```

Db      |||
181 GATCTGACTGAGCTGCTGCTGACCACTGCTCGAAGAAAGACAGCTGCTTC 240
QY      936 CTAGTACAGAGAAAGAGGCTTGGGCTTGGAGAGAGAGATCTAGCTGGCACT 995
        |||
Db      241 CTAGTACAGAGAAAGAGGCTTGGGCTTGGAGAGAGAGATCTAGCTGGCACT 300
QY      996 TCACATCCGCTCCTCTGCTCAGAGGCTGAGAGGAGGAGCTGGAATACCCCTAGTAT 1055
        |||
Db      301 TCACATCCGCTCCTCTGCTCAGAGGCTGAGAGGAGGAGCTGGAATACCCCTAGTAT 360
QY      1056 GGAATGAGAGGCTTGGGCTTGGAGAGAGATCTAGCTGGGCTGCT 1115
        |||
Db      361 GGAATGAGAGGCTTGGGCTTGGAGAGAGATCTAGCTGGGCTGCT 420
QY      1116 TCCTTCTCTGATACGGGAGAGACCC 1142
        |||
Db      421 TCCTTCTCTGATACGGGAGAGACCC 447

```

RESULT 15

AAV86113/C

ID AAV86113 standard; cDNA; 450 BP.

XX AC AAV86113;

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

XX AC AAV86113; (first entry)

```

SQ      Sequence 450 BP; 86 A; 152 C; 120 G; 92 T; 0 other:
Query Match      36.4%; Score 433.8; DB 20; Length 450;
Best Local Similarity 98.4%; Pred. No. 1,2e-100;
Matches 438; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY      550 CGAGGGGCCACCGAGCGCTAGCAGTGGGGGAGACGCGCCAGACCTGGGCTTGGACCT 609
        |||
Db      445 CGAGGGGCCAAGAGAGCGCTAGCAGTGGGGGAGACGCGCCAGACCTGGGCTTGGACCT 386
QY      610 GCTGGGCCCAACGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 669
        |||
Db      385 GCTGGGCCCAACGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 326
QY      670 TGACACCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 729
        |||
Db      325 TGACACCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 266
QY      730 CCTGAGAGTGGGCGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 789
        |||
Db      265 CCTGAGAGTGGGCGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 206
QY      790 CCTGAGAGTGGGCGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 849
        |||
Db      205 CCTGAGAGTGGGCGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 146
QY      850 AGCCCTGCTGCTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 909
        |||
Db      145 AGCCCTGCTGCTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 86
QY      910 TCCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 969
        |||
Db      85 TCCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 26
QY      970 GAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 994
        |||
Db      25 GAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1

```

Search completed: August 25, 2003, 13:51:51
Job time : 374 secs

Claim 1; Page 125; 633pp; English.

This sequence represents an expressed sequence tag (EST), and is a polynucleotide of the invention. The polynucleotides of the invention are all secreted EST sequences isolated from a variety of human tissue sources. The EST sequences and proteins encoded by them are predicted to have useful biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional activity, immune stimulating or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, tumour inhibition activity. The EST sequences are also stated to be useful for gene therapy.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 12:19:01 ; Search time 4428 Seconds

(without alignments)
11012.709 Million cell updates/sec

Title: US-09-701-675A-8

Perfect score: 1192
Sequence: 1 gtagccgagtcagactg.....gggaatctcaaaaaaaaaa 1192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pin: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1183.4	99.3	2113	6 BD127703	BD127703 Primer fo
2	1183.4	99.3	2113	9 AK074652	AK074652 Homo sapi
3	1159.4	97.3	1179	9 BC002670	BC002670 Homo sapi
4	1159	97.2	1218	9 AF117959	AF117959 Homo sapi
5	1151.2	96.6	1182	6 BD119394	BD119394 Homo sapi
6	1130.4	94.8	1178	9 AF366402	AF366402 Extended
7	1048	87.9	5917	9 AF130860	AF130860 Homo sapi
8	1046.4	87.8	160643	9 AC010271	AC010271 Homo sapi
9	777.8	65.3	5746	6 AX472956	AX472956 Sequence
10	773.8	61.9	978	9 HSA323696	HSA323696 Homo sapi
11	675.6	56.7	1247	10 AF366401	AF366401 Mus muscu
12	667.4	56.0	5869	6 AX537459	AX537459 Sequence
13	640.8	53.8	1165	10 BC016077	BC016077 Mus muscu
14	637.2	53.5	696	6 BD124736	BD124736 Primer fo
15	637.2	53.5	696	6 BD126760	BD126760 Primer fo
16	588	49.3	145347	2 AC073821	AC073821 Mus muscu
17	573.4	48.1	877	10 AF218291	AF218291 Mus muscu
18	569.6	47.8	201377	2 AC073767	AC073767 Mus muscu
19	553.6	46.4	257945	2 AC112313	AC112313 Rattus no
20	553.6	46.4	290929	2 AC118914	AC118914 Rattus no
21	532.4	44.7	164652	2 AC021625	AC021625 Homo sapi
22	501	42.0	701	10 AF366400	AF366400 Mus muscu
23	433.8	36.4	450	6 BD058236	BD058236 Secreted
24	419.8	35.2	745	9 HSA325585	HSA325585 Homo sapi
25	391.6	32.9	476	6 BD109590	BD109590 EST and e
26	330	27.7	256418	2 AC115309	AC115309 Rattus no
27	293.4	24.6	301	6 BD076276	BD076276 5' EST of
28	237.6	19.9	536	6 AX283999	AX283999 Sequence
29	227.6	19.1	68118	2 AC011113	AC011113 Homo sapi
30	172.8	14.5	986	9 HSA233697	HSA233697 Homo sapi
31	69.6	5.8	125020	9 AF429315	AF429315 Homo sapi
32	67.2	5.5	173048	5 AL772154	AL772154 Zebrafish
33	66	5.5	173048	5 AL772154	AL772154 Zebrafish
34	63.4	5.3	127051	2 AC114895	AC114895 Tetradon
35	59.8	5.0	134408	2 AC139628	AC139628 Takifugu
36	58.6	4.9	2000	6 AX655393	AX655393 Sequence
37	50.4	4.2	314100	1 SC039106	SC039106 Streptomy
38	47.4	4.0	2000	6 AX655393	AX655393 Sequence
39	47	3.9	1885	9 AF514298	AF514298 Homo sapi
40	47	3.9	1949	9 AF093445	AF093445 Homo sapi
41	47	3.9	138999	9 AC090426	AC090426 Homo sapi
42	47	3.9	158766	2 AC015851	AC015851 Homo sapi
43	47	3.9	166234	9 AC018629	AC018629 Homo sapi
44	47	3.9	173440	2 AC080112	AC080112 Homo sapi
45	47	3.9	188574	2 AC131063	AC131063 Homo sapi

ALIGNMENTS

RESULT 1
BD127703
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2113)
REFERENCE
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.


```

FEATURES
  source
    location/Qualifiers
      1..2113
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="MAMMA1000403"
      /tissue_type="mammary gland"
      /clone_lib="MAMMA1"
      /note="cloning vector: pME18SFL3"
BASE COUNT      453 a      561 c      578 g      521 t
ORIGIN
Query Match      99.3%; Score 1183.4; DB 9; Length 2113;
Best Local Similarity 99.9%; Pred. No. 2.3e-251;
Matches 1184; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGAGCGGAGTAGAAGTGCCTGCGAGCCAGGCGGGGGTTCCGGAGGACACCA 60
   |||
Db 21 GTAGCGCGAGTAGAAGTGCCTGCGAGCCAGGCGGGGGTTCCGGAGGACACCA 80
   |||

QY 61 CAAGCATGCTGCGCGCGCGCTTCCGTGATTTGGTTGGGTGGCTACTCTTCTCTGA 120
   |||
Db 81 CAAGCATGCTGCGCGCGCGCTTCCGTGATTTGGTTGGGTGGCTACTCTTCTCTGA 140
   |||

QY 121 TTGGCCGCTAGTAGACAAGTCTGAGCAAGGCTCGAAGCGGAACGGAGAGAGA 180
   |||
Db 141 TTGGCCGCTAGTAGACAAGTCTGAGCAAGGCTCGAAGCGGAACGGAGAGAGA 200
   |||

QY 181 GGAGAGGAACCTCTGGAGTGCATCTCGTGGTGGCTAGATCTTGGGCGACGACGGTGC 240
   |||
Db 201 GGAGAGGAACCTCTGGAGTGCATCTCGTGGTGGCTAGATCTTGGGCGACGACGGTGC 260
   |||

QY 241 ACAGGACCCCGCGCGCGCTGAGCTCCCTCTTTGACCTCTCAAGTCAAGTCCA 300
   |||
Db 261 ACAGGACCCCGCGCGCGCTGAGCTCCCTCTTTGACCTCTCAAGTCAAGTCCA 320
   |||

QY 301 CCACAGCTCTGACAGAGTAGAGCCGACCTGCGGCACTGTGTGCTGTGAACACTCT 360
   |||
Db 321 CCACAGCTCTGACAGAGTAGAGCCGACCTGCGGCACTGTGTGCTGTGAACACTCT 380
   |||

QY 361 GCGGCGCATCGAGGCTCCATGAGCAGCCGCGGCTGCCCTGCGCACTGTGCTTACGCCAC 420
   |||
Db 381 GCGGCGCATCGAGGCTCCATGAGCAGCCGCGGCTGCCCTGCGCACTGTGCTTACGCCAC 440
   |||

QY 421 TCGAGCCCGCAGTGTGCTGACAACTTACTGCGCAAGCTCGAGAGCTGCCCTTACAGCTTC 480
   |||
Db 441 TCGAGCCCGCAGTGTGCTGACAACTTACTGCGCAAGCTCGAGAGCTGCCCTTACAGCTTC 500
   |||

QY 481 CATGCGCAGCTCTCTGGAGAGCTCGACCAATTGAGGCTGTAGAGCTCCCAACC 540
   |||
Db 501 CATGCGCAGCTCTCTGGAGAGCTCGACCAATTGAGGCTGTAGAGCTCCCAACC 560
   |||

QY 541 CTTGGCAGAGAGAGGCGCACAGAGCGCTGATCGGGGGAGAGCGCCCAAGCTTGAGTGC 600
   |||
Db 561 CTTGGCAGAGAGAGGCGCACAGAGCGCTGATCGGGGGAGAGCGCCCAAGCTTGAGTGC 620
   |||

QY 601 CTTGGAAGCTGTGGGCGCACAGCACTGCTGTCTACTGAGAGATGGGCTTGAGGGCTGTT 660
   |||
Db 621 CTTGGAAGCTGTGGGCGCACAGCACTGCTGTCTACTGAGAGATGGGCTTGAGGGCTGTT 680
   |||

QY 661 TCGAGATATTTGACACTCTTATGTATGACAAATGAACTTTGGGCGACAGCCTTGAGGGCCT 720
   |||
Db 681 TCGAGATATTTGACACTCTTATGTATGACAAATGAACTTTGGGCGACAGCCTTGAGGGCCT 740
   |||

QY 721 CAATCCAGGCGCTGAGATGAGGCGGGCAAGAGGAAGCTCCGAGACTGGAGAGGCCCA 780
   |||
Db 741 CAATCCAGGCGCTGAGATGAGGCGGGCAAGAGGAAGCTCCGAGACTGGAGAGGCCCA 800
   |||

QY 781 ATTGGAATCTATGATGATGTGCTGTGGGCAACAGGAGCTTGAGAGCGACCGCGGGGCG 840
   |||
Db 801 ATTGGAATCTATGATGATGTGCTGTGGGCAACAGGAGCTTGAGAGCGACCGCGGGGCG 860
   |||

QY 841 AGGCGCGTAGAGCCCTGCTGCTGGAATGATTGCTGTGTTGATCTGAACCTGCTGCTG 900
   |||

```

```

Db 861 AGGCGCGTAGAGCCCTGCTGGAATGATTGTCTGATCTGAACAGAGCTCTGAGCTG 920
   |||
QY 901 GACCAACTGTCCTCGAAAAGACACACACTGGCTCCCTAGTACAGATAAAGGGCTGGCG 960
   |||
Db 921 GACCAACTGTCCTCGAAAAGACACACACTGGCTCCCTAGTACAGATAAAGGGCTGGCG 980
   |||
QY 961 CACTTTGAGAGACAGAAATCTAGTCTGGGCAACTTCACATCCGCTCCTGTCTCAGGG 1020
   |||
Db 981 CACTTTGAGAGACAGAAATCTAGTCTGGGCAACTTCACATCCGCTCCTGTCTCAGGG 1040
   |||

QY 1021 CTGGCAGGGGAGACCTGGAATTAACCCCTTAGTATGGAATGACAGGCTGTGAGGACT 1080
   |||
Db 1041 CTGGCAGGGGAGACCTGGAATTAACCCCTTAGTATGGAATGACAGGCTGTGAGGACT 1100
   |||

QY 1081 GAATTCCTGGGCGCTGAGGATAGCTGTGGGCTGCTCTCTGTGATGAGGGAAGAGACC 1140
   |||
Db 1101 GAATTCCTGGGCGCTGAGGATAGCTGTGGGCTGCTCTCTGTGATGAGGGAAGAGACC 1160
   |||

QY 1141 CCATCAGATTTTTCAAATTAAAGCCAGTCTGGAAATCTCAAA 1185
   |||
Db 1161 CCATCAGATTTTTCAAATTAAAGCCAGTCTGGAAATCTCAAA 1205
   |||

```

```

RESULT 3
BC002670      1179 bp      mRNA      linear      PRI 12-JUL-2001
LOCUS
DEFINITION
  Homo sapiens, CDK4-binding protein p34SEI1, clone MGC:3465
ACCESSION
  BC002670
VERSION
  BC002670.1 GI:12803668
KEYWORDS
  MGC.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

REFERENCE
  1 (bases 1 to 1179)
  Strausberg, R.
  Direct Submision
  Submitted (05-FEB-2001) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
  USA

```

```

REMARK
  NIH-MGC Project URL: http://mgc.ncl.nih.gov
  Contact: MGC help desk
  Email: cgabs-r@mail.nih.gov
  Tissue Procurement: ATCC

```

```

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
  Sequencing Center (NISC),
  Gaithersburg, Maryland;
  Web site: http://www.nisc.nih.gov/

```

```

Contact: nisc_mgc@ncl.nih.gov
  Shevchenko, Y., Wehtherby, K.D., Beckstrom-Sternberg, S.M.,
  Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
  Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
  Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
  McDowell, J., Pearson, R., Snyder, B., Stantirpop, S., Thomas, P.D.,
  Tlionson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
  Zhang, L.-H. and Green, E.D.

```

```

Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LNL at: http://lmln1.gov
  Series: IRAL Plate: 12 Row: i Column: 23
  This clone was selected for full length sequencing because it
  passed the following selection criteria: matched mRNA gi: 7019524.

```

```

FEATURES
  source
    1..1179
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="locustid:29950"
    /db_xref="taxon:9606"

```

```

CDS
      118..828
      /codon_start=1
      /product="CDK4-binding protein p34SEI1"
      /protein_id="AAH02670.1"
      /db_xref="GI:12803669"
      /translation="MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAPPAVASS
      LFDLSVLKHSLSQSEPDRLHVLVYTLRIQASMAPAALPVPSPAPASVDN
      LIASDPAALSASMSLSLEDLSHIEGLSQAPPLADEPSPGSIQGAAPSLGALDLGP
      AVGCLLDGLESLEPDIDTYSMDNELMVAPESEGLKPGEDPGEKEAPELEAIDLGP
      MDVLVGTQLELRPPGPR"
BASE COUNT      243 a      343 c      368 g      225 t
ORIGIN
Query Match      97.3%; Score 1159.4; DB 9; Length 1179;
Best Local Similarity 99.9%; Pred. No. 5e-246;
Matches 1160; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
      32 CCAGGCGGGGTTTCCGGAGACAGCAACAAGCATGTCGCCGCCGTTTCTGAT 91
      10 CCAGGCGGGGTTTCCGGAGACAGCAACAAGCATGTCGCCGCCGTTTCTGAT 69
      92 GGTGTGGGTGGCTACCTCTTCTGATTGGCCGCTAGTAGAAGATGCTGACAG 151
      70 GGTGTGGGTGGCTACCTCTTCTGATTGGCCGCTAGTAGAAGATGCTGACAG 129
      152 GGTGGAAGCGGAAGGAGAGAGAGAGGAAGAAAGCACTGCGAGTGCACCTCGG 211
      130 GGTGGAAGCGGAAGGAGAGAGAGAGGAAGAAAGCACTGCGAGTGCACCTCGG 189
      212 TGGCTAGATCTTGGCCACAGCGGTGGACAGAGGACCCCGCGCTCTAGCTCC 271
      190 TGGCTAGATCTTGGCCACAGCGGTGGACAGAGGACCCCGCGCTCTAGCTCC 249
      272 CTCTTTGACCTCTCAGTGTCTCAAGTCCACACAGCTGACAGAGTGAAGCCGAG 331
      250 CTCTTTGACCTCTCAGTGTCTCAAGTCCACACAGCTGACAGAGTGAAGCCGAG 309
      332 CGGACCGGTCGTGTGTGTGAACACTGTGGGGGACATCCAGGCGTCCATGGACACCCG 391
      310 CGGACCGGTCGTGTGTGTGAACACTGTGGGGGACATCCAGGCGTCCATGGACACCCG 369
      392 GCTGCCCTGACCACTGTGTGTGTGAACACTGTGGGGGACATCCAGGCGTCCATGG 451
      370 GCTGCCCTGACCACTGTGTGTGTGAACACTGTGGGGGACATCCAGGCGTCCATGG 429
      452 GCAAGCTGGAGCGTGTGTGTGTGAACACTGTGGGGGACATCCAGGCGTCCATGG 511
      430 GCAAGCTGGAGCGTGTGTGTGTGAACACTGTGGGGGACATCCAGGCGTCCATGG 489
      512 ATTGAGGCGCTGAGTCAAGGCTCCCAACCTTGGAGAGAGAGGGGCCACAGGCCGTAG 571
      490 ATTGAGGCGCTGAGTCAAGGCTCCCAACCTTGGAGAGAGAGGGGCCACAGGCCGTAG 549
      572 ATCGGGGAGAGCGGCCAGCTGTGGTGTGTGAACCTGTGGGGGCCACAGCTGGCTGT 631
      550 ATCGGGGAGAGCGGCCAGCTGTGGTGTGTGAACCTGTGGGGGCCACAGCTGGCTGT 609
      632 CTACTGGAGATGGGCTTGAAGGCGCTGTTGAGATATATGACACTCTATATGACAA 691
      610 CTACTGGAGATGGGCTTGAAGGCGCTGTTGAGATATATGACACTCTATATGACAA 669
      692 GAATTTGGGACAGCCTCTGAGGGGCTCAACACAGGCGCTGAGAGATGGGCGGGAG 751
      670 GAATTTGGGACAGCCTCTGAGGGGCTCAACACAGGCGCTGAGAGATGGGCGGGAG 729
      752 GAGGAAGCTCCGAGCTGAGAGAGAGCGCAATTTGACTACCTCATGATGTCTGTGGG 811
      730 GAGGAAGCTCCGAGCTGAGAGAGAGCGCAATTTGACTACCTCATGATGTCTGTGGG 789

```

```

      812 ACACAGGACATGGAGCGAGCCCGCGGCGCCAGAGCGCTGAGCCCTGCTGGAATGGTTG 871
      790 ACACAGGACATGGAGCGAGCCCGCGGCGCCAGAGCGCTGAGCCCTGCTGGAATGGTTG 849
      872 TCTGTATCTGAATGACGACCTGCTGCTGAGACACACTCTCTCCGAAAAGACACAGCTGGC 931
      850 TCTGTATCTGAATGACGACCTGCTGCTGAGACACACTCTCTCCGAAAAGACACAGCTGGC 909
      932 TTCCCTAGTACAGAGAACAGGCGCTTGGGCACTTTGGAGAGACAGAACTAGTCTGGGC 991
      910 TTCCCTAGTACAGAGAACAGGCGCTTGGGCACTTTGGAGAGACAGAACTAGTCTGGGC 969
      992 AACTTCACATCCGTCCTCTCTCTGAGGCTGCGACAGGGGAGACCTGAATTAACCCCTAG 1051
      970 AACTTCACATCCGTCCTCTCTCTGAGGCTGCGACAGGGGAGACCTGAATTAACCCCTAG 1029
      1052 TGATGGAATGACAGGCTGCTGGTGGGAGACTGAATTCCTGCGCCCTGAGTACTGGG 1111
      1030 TGATGGAATGACAGGCTGCTGGTGGGAGACTGAATTCCTGCGCCCTGAGTACTGGG 1089
      1112 CTGTTCTCTCTCTGATACGGAAGAGACCCCAATCAGATTTTTCAAATTAAGCCAGTCC 1171
      1090 CTGTTCTCTCTCTGATACGGAAGAGACCCCAATCAGATTTTTCAAATTAAGCCAGTCC 1149
      1172 TGGCAATCTCAAAAAAAA 1192
      1150 TGGCAATCTCAAAAAAAA 1170
      RESULT 4
      AF117959
      LOCUS      Homo sapiens CDK4-binding protein p34SEI1 (SEI1) mRNA, complete cds.
      DEFINITION
      VERSION      AF117959.1 GI:6434875
      KEYWORDS
      SOURCE
      ORGANISM      Homo sapiens (human)
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
      REFERENCE
      AUTHORS
      TITLE
      JOURNAL
      PUBMED
      MEDLINE
      JOURNAL
      10580009
      2 (bases 1 to 1218)
      Hara,E. and Sugimoto,M.
      Direct Submission
      Submitted (05-JAN-1999) Cell Cycle Laboratory, Paterson Institute
      for Cancer Research, Wilmslow Road, Manchester M20 4BX, UK
      FEATURES
      source
      1..1218
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /chromosome="19"
      /map="19q13.1-q13.2"
      1..1218
      /gene="SEI1"
      131..841
      /gene="SEI1"
      /codon_start=1
      /product="CDK4-binding protein p34SEI1"
      /protein_id="AAH08349.1"
      /db_xref="GI:6434876"
      /translation="MLSKGLKRRKEEKEEPLAVDSWMLDPGHAAVAQAPPAVASS
      LFDLSVLKHSLSQSEPDRLHVLVYTLRIQASMAPAALPVPSPAPASVDN
      LIASDPAALSASMSLSLEDLSHIEGLSQAPPLADEPSPGSIQGAAPSLGALDLGP

```

```

ATGCLLDGLEGLEFEDIDTSMTDNEIMAPASBGLKRPEDPGKEBAPBLDAEIDLTY
MDLVETQLEPBPGR*
BASE COUNT      274 a      347 c      369 g      228 t
ORIGIN
Query Match      97.2%; Score 1159; DB 9; Length 1218;
Best Local Similarity 99.5%; Pred. No. 6.1e-246;
Matches 1173; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 14 AGAAGTCGCTCTGCGACGACGCGGGGTTCCGGAGAGACGCCAACAAGCATCTGC 73
Db 6 AGAAGTCGCTCTGCGACGACGCGGGGTTCCGGAGAGACGCCAACAAGCATCTGC 64
QY 74 CGCGCGCGTTCTCTGATTGGTGTGGGGTACCTCTTCGTTGATTGGCCGCTAGTG 133
Db 65 CGCGCGCGTTCTCTGATTGGTGTGGGGTACCTCTTCGTTGATTGGCCGCTAGTG 124
QY 134 AGCAAGATCTGAGCAAGGGGTCTGAAGCGGAAACGGAGAGAGAGAGAAAGAACT 193
Db 125 AGCAAGATCTGAGCAAGGGGTCTGAAGCGGAAACGGAGAGAGAGAGAAAGAACT 184
QY 194 CTGGCAGTCACTCCGTTGAGTCTGAGTCTGAGCAGCGGTTGGCACAGGACCCCG 253
Db 185 CTGGCAGTCACTCCGTTGAGTCTGAGTCTGAGCAGCGGTTGGCACAGGACCCCG 244
QY 254 GCGTGGCGCTTAGCTCCCTCTTTGACCTCTGAGTCTGAGCTCCACAGCCTGACG 313
Db 245 GCGTGGCGCTTAGCTCCCTCTTTGACCTCTGAGTCTGAGCTCCACAGCCTGACG 304
QY 314 CAGAGTGAACCGGACTTGGGACGACCTGGTGTGCTGTAACACTTGGCGGATCCAG 373
Db 305 CAGAGTGAACCGGACTTGGGACGACCTGGTGTGCTGTAACACTTGGCGGATCCAG 364
QY 374 GCGTCCATGAGCAACCGCGGCTCCCTGACCACTGTGCTAGCCCACTCAGCCCGCAGT 433
Db 365 GCGTCCATGAGCAACCGCGGCTCCCTGACCACTGTGCTAGCCCACTCAGCCCGCAGT 424
QY 434 GTGGCTGACACTTACTGAGCAAGCTGCGAGCTGCCCTTTGACCTGATGGCCAGCTC 493
Db 425 GTGGCTGACACTTACTGAGCAAGCTGCGAGCTGCCCTTTGACCTGATGGCCAGCTC 484
QY 494 CTGGAGGACCTGAGCAACTTGAAGGGCGTGAAGGGTCCCAACCCCTTGGACAGAG 553
Db 485 CTGGAGGACCTGAGCAACTTGAAGGGCGTGAAGGGTCCCAACCCCTTGGACAGAG 544
QY 554 GGGCCACAGCGCGCTGATGATCGGGGAGAGCGCCAGCCTGGTGCCTTTGAGACCTG 613
Db 545 GGGCCACAGCGCGCTGATGATCGGGGAGAGCGCCAGCCTGGTGCCTTTGAGACCTG 604
QY 614 GGGCCACAGCGCGCTGATGATCGGGGAGAGCGCCAGCCTGGTGCCTTTGAGACCTG 673
Db 605 GGGCCACAGCGCGCTGATGATCGGGGAGAGCGCCAGCCTGGTGCCTTTGAGACCTG 664
QY 674 ACCTGATGATGACAAATGACTTTGGGACAGCGCTGAGAGGCTCAAAACAGGCTC 733
Db 665 ACCTGATGATGACAAATGACTTTGGGACAGCGCTGAGAGGCTCAAAACAGGCTC 724
QY 734 GAGATGAGCGCGGCAAGAGAGAGCTCGGAGCTGAGAGAGCGCAATTGGACTACCTC 793
Db 725 GAGATGAGCGCGGCAAGAGAGAGCTCGGAGCTGAGAGAGCGCAATTGGACTACCTC 784
QY 794 ATGATGATGCTGTGGGACACAGGCACTGAGCGACCGCGGGGCGCAGGGCGCTGAGCC 853
Db 785 ATGATGATGCTGTGGGACACAGGCACTGAGCGACCGCGGGGCGCAGGGCGCTGAGCC 844
QY 854 CTGCGTCTGAGTGTGCTGTGCTGATCTGAATGAGCTGCTGGAGCAACATGCTCT 913
Db 845 CTGCGTCTGAGTGTGCTGTGCTGATCTGAATGAGCTGCTGGAGCAACATGCTCT 904
QY 914 CGAAAGACACAGCTGCTCCCTAGTACAGAAACAGGGCTTGGGCGCTTTGGAGAGA 973
Db 905 CGAAAGACACAGCTGCTCCCTAGTACAGAAACAGGGCTTGGGCGCTTTGGAGAGA 964

```

```

QY 974 CAGAATCTAGCTCCGGGCAACTTCACATCCGTCCTCTCTCAGGCTGGCAGGGGAG 1033
Db 965 CAGAATCTAGCTCCGGGCAACTTCACATCCGTCCTCTCTCAGGCTGGCAGGGGAG 1024
QY 1034 CCTGCAATTAACCCCTAGTGAATGATGACAGGGTCTGGTGGGAGTGAATTCCTGGCC 1093
Db 1025 CCTGCAATTAACCCCTAGTGAATGATGACAGGGTCTGGTGGGAGTGAATTCCTGGCC 1084
QY 1094 CTGGGGTCAATTAACCTGGGCTGTCTCTCTGATACGGGAGAGACCCCAATCAGATTTT 1153
Db 1085 CTGGGGTCAATTAACCTGGGCTGTCTCTCTGATACGGGAGAGACCCCAATCAGATTTT 1144
QY 1154 TCAAATTAAGCCAGTCTGAGAAATCTCAAAAAAAAAA 1192
Db 1145 TCAAATTAAGCCAGTCTGAGAAATCTCAAAAAAAAAA 1183

RESULT 5
BD139394      1182 bp      DNA      linear      PAT 18-SEP-2002
LOCUS      Extended cDNA of secretory protein.
DEFINITION      BD139394
ACCESSION      BD139394.1 GI:22324339
VERSION      JP 2002508182-A/146.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 1182)
AUTHORS      Bouguetel, L., Ducleert, A. and Edwards, J. B. D. M.
TITLE      Extended cDNA of secretory protein
JOURNAL      Patent: JP 2002508182-A 146 19-MAR-2002;
GENSET
OS      Homo sapiens (human)
PN      JP 2002508182-A/146
PD      19-MAR-2002
PF      17-DEC-1998 JP 2000539136
PR      17-DEC-1997 US 60/069957, 09-FEB-1998 US 60/074121 PR
13-APR-1998 US 60/081563, 10-AUG-1998 US 60/096116 PI LYDIE
BOUGUETEL, L., DUCLEERT, A., JEAN BAPTISTE DUMAS MIENE PI EDMARDS
PC      C12N1/21, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC
PC      C12N5/10, C12P21/02, C12Q1/68, C12N15/00, C12N5/00, C12N15/00 CC
Von Heljne matrix
CC      score 4.6999980926514
FH      Key      Location/Qualifiers
FT      CDS      128..835
FT      sig_peptide      128..220
FT      polyA_signal      1145..1150
FT      polyA_site      1170..1181.
FEATURES
source      1..1182
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT      239 a      348 c      357 g      227 t      11 others
ORIGIN
Query Match      96.6%; Score 1151.2; DB 6; Length 1182;
Best Local Similarity 98.6%; Pred. No. 3.2e-244;
Matches 1162; Conservative 8; Mismatches 8; Indels 1; Gaps 1;

QY 14 AGAAGTCGCTCTGCGACGACGCGGGGTTCCGGAGAGACGCCAACAAGGATGCTGC 73
Db 2 AGAAGTCGCTCTGCGACGACGCGGGGTTCCGGAGAGACGCCAACAAGGATGCTGC 61
QY 74 CGCGCGCGTTCTCTGATTGGTGTGGGGTACCTCTTCGTTGATTGGCCGCTAGTG 133
Db 62 CGCGCGCGTTCTCTGATTGGTGTGGGGTACCTCTTCGTTGATTGGCCGCTAGTG 121
QY 134 AGCAAGATCTGAGCAAGGGGTCTGAAGCGGAAACGGAGAGAGAGAGAAAGAACT 193

```

Db	122	AGCAAGATGCTGAGCAAGGGTCTGAAGCGGAACGGGAGAGAGAGAGAGAAAGAACT	181
OY	194	CTGGCAGTGCAGTCCCTGGGTGCTAGATCCTGGCCACGACAGCGGTGGCACAGGCCG	253
Db	182	CTGGCAGTGCAGTCCCTGGGTGCTAGATCCTGGCCACGACAGCGGTGGCACAGGCCG	241
OY	254	GCCGGGCTCTAGCTCCCTCTTTGACCTCTAGTGTCTAAGCTCCACCACAGCCTGCAG	313
Db	242	GCCGGGCTCTAGCTCCCTCTTTGACCTCTAGTGTCTAAGCTCCACCACAGCCTGCAG	301
OY	314	CAGAGTGAACCGGACCTCGGAGACCTGGTGGTGGTGGTGAACATCTGGCGGGCATCCAG	373
Db	302	VRRAATBAGCCGGACCTCGGAGACCTGGTGGTGGTGGTGAACATCTGGCGGGCATCCAG	361
OY	374	GCCTCCATGGACACCCCGGGCTGCCTGCCACCTGTGCTTAGGCCACCTGCAGCCCAAT	433
Db	362	GCCTCCATGGACACCCCGGGCTGCCTGCCACCTGTGCTTAGGCCACCTGCAGCCCAAT	421
OY	434	GTGGGTGACAACTTACTGGCAAGCTCGGACGCTGGCTTTTCAAGCTCCATGGCCAGCTC	493
Db	422	GTGGGTGACAACTTACTGGCAAGCTCGGACGCTGGCTTTTCAAGCTCCATGGCCAGCTC	481
OY	494	CTGGAGGACCTCAGGCACATTTGAGGGCCGTAAGTCAAGGCTCCCAACCCCTGGCACAACG	553
Db	482	CTGGAGGACCTCAGGCACATTTGAGGGCCGTAAGTCAAGGCTCCCAACCCCTGGCACAACG	541
OY	554	GGGCCACACAGGCGCTGAGATCGGGGGAGACAGGCCACCTGGGTGCTTTGGACCTGCTG	613
Db	542	GGGCCACACAGGCGCTGAGATCGGGGGAGACAGGCCACCTGGGTGCTTTGGACCTGCTG	601
OY	614	GGCCACGACACTGGGTGTCTACTGGACGATGGGCTTTGAGGGCTTTTGAAGATTTGAC	673
Db	602	GGCCACGACACTGGGTGTCTACTGGACGATGGGCTTTGAGGGCTTTTGAAGATTTGAC	661
OY	674	ACCTTATGTATGACAAATGAATTTGGGCACACAGGCTCTGAGAGGGCTCAAAACAGAGCCT	733
Db	662	ACCTTATGTATGACAAATGAATTTGGGCACACAGGCTCTGAGAGGGCTCAAAACAGAGCCT	721
OY	734	GAGGATGGGCGCGGGCAGAGGAACTCCGACCTGGACGTGACAGAGCCGAAATTTGGACTACCTC	793
Db	722	GAGGATGGGCGCGGGCAGAGGAACTCCGACCTGGACGTGACAGAGCCGAAATTTGGACTACCTC	781
OY	794	ATGGAATGGCTGGTGGGACACACAGGCACATGGAGCGACCCGGGGGCCAGGGGCTGAGGC	853
Db	782	ATGGAATGGCTGGTGGGACACACAGGCACATGGAGCGACCCGGGGGCCAGGGGCTGAGGC	841
OY	854	CTCGGTGCGAATGGTGTCTGTGTATCTGAACCTGAGCCTGCGGGCTGGACCAACTGTCT	913
Db	842	CTCGGTGCGAATGGTGTCTGTGTATCTGAACCTGAGCCTGCTGGCTGGACCAACTGTCT	901
OY	914	CGAAAAGACACAGCTGGCTCCCTAGTACAGAGAAACAGGGCTTGGGCCACTTTGGAGAGA	973
Db	902	CGAAAAGACACAGCTGGCTCCCTAGTACAGAGAAACAGGGCTTGGGCCACTTTGGAGAGA	961
OY	974	CAGAAATCTAGTCTGGGCACTTCAACATCGCTCCTGTCTCAAGGCTGGCAGGGGGAG	1033
Db	962	CAGAAATCTAGTCTGGGCACTTCAACATCGCTCCTGTCTCAAGGCTGGCAGGGGGAG	1021
OY	1034	CCCTGAATTTACCCCTAGTGAATGGAATGACAGGGGCTGGTGGGGACCTAAATTTCCCTGGCC	1093
Db	1022	CCCTGAATTTACCCCTAGTGAATGGAATGACAGGGGCTGGTGGGGACCTAAATTTCCCTGGCC	1081
OY	1094	CTGGGGTCATAGCTTGGGCTGTCTCTTCTGATACGGGAGAGAGACCCCAATCAGATTTT	1155
Db	1082	CTGGGGTCATAGCTTGGGCTGTCTCTTCTGATACGGGAGAGAGACCCCAATCAGATTTT	1143
OY	1154	TCAAAATTAAGCCAGTCTGGGAAATCTCAAAAAAAA 1192	
Db	1141	TCAAAATTAAGCCAGTCTGGGAAATCTCAAAAAAAA 1179	

LOCUS	AF366402	1178 bp	mRNA	linear	PRI 13-MAY-2001
DEFINITION	Homo sapiens TRIP-Brl mRNA, complete cds.				
ACCESSION	AF366402				
VERSION	AF366402.1	GI:14029833			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Hsu,S.I., Yang,C.M., SIm,K.G., Hentschel,D.M., O'Leary,E. and Bonventre,J.V.				
TITLE	TRIP-Brl: a novel family of PHD zinc finger- and bromodomain-interacting proteins that regulate the transcriptional activity of E2F-1/DP-1				
JOURNAL	EMBO J. 20 (9), 2273-2285 (2001)				
MEDLINE	21231173				
PubMed	11331592				
REFERENCE	2 (bases 1 to 1178)				
AUTHORS	Hsu,S.I. and Bonventre,J.V.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-Mar-2001) Dept. of Medicine, National University Hospital, 5 lower Kent Ridge Road 119074, Singapore				
FEATURES	Location/Qualifiers				
source	1..1178				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="dbEST:w52056"				
	/db_xref="taxon:9606"				
	/chromosome="19"				
	/map="19q13"				
	112..822				
	/function="acts at E2F-responsive promoters to integrate signals provided by PHD- and/or bromodomain-containing transcription factors"				
	/function="a CDK4-binding protein that renders the activity of cyclin D/CDK4 resistant to the inhibitory effects of p16(INK4a)"				
	/note="hTRIP-Brl; transcriptional regulator interacting with the PHD-bromodomain; hTRIP-Brl(SBI-1)"				
	/codon_start=1				
	/product="TRIP-Brl"				
	/protein_id="AAK52831.1"				
	/db_xref="GI:14029834"				
	/translation="MLSGLKRRKEEREKEPLAVDSWMLDPGHAAVAQAPPAVASSSS LFDLSVLKHSLOOSEPDRLHLVYVWTLERIQASMAPALPPVSPPAQSVADN LLAASDPAASMASLIDELSHIGLSQAPPLADEGPPGSGGAAPSGALDLRR ATGCLDDELGLELFEDLIDTSYNDLMNAPASEGLKPPEDSGKEAPELDEALDYIL MDVAVGTALERPPEGR"				
BASE COUNT	245 a	343 c	365 g	225 t	
ORIGIN					
Query Match	94.8%	Score 1130.4;	DB 9;	Length 1178;	
Best Local Similarity	99.1%	Pred. No. 1.3e-239;			
Matches 1158;	Conservative	0;	Mismatches	6;	Indels 4; Gaps 2;
26	CGCGACCCAGGACGGCGGTTTCCGAGAGACACCAACAAGCATGCGCGCGCGTTTC	85			
Db	1 CGCGACCCAGGCGAG--TTTCCGAGAGCGACCAACAAGCATGCGCGCGCGTTTC	57			
QY	86 CTGATTGGTTGGGTGGGTACTCTTCTGTTCTGATTGGCGCTAGTGACMAATGCTG	145			
Db	58 CTGATTGGTTGGGTGGGTGGGTACTCTTCTGTTCTGATTGGCGCTAGTGACMAATGCTG	117			
QY	146 AGCAAGGTTCTGAAGCGGAAACGGAGAGAGAGAGAGAGAGAACTCTGGCAGTGCAC	205			
Db	118 AGCAAGGTTCTGAAGCGGAAACGGAGAGAGAGAGAGAGAGAACTCTGGCAGTGCAC	177			
QY	206 TCCGTTGGTGGCAGAGATCCCTGAGCAGCGAGCGGTGGCACAGGACCCCGGGCGTGGCT	265			
Db	178 TCCGTTGGTGGCAGATCCCTGAGCAGCGAGCGGTGGCACAGGACCCCGGGCGTGGCT	237			
QY	266 AGCTCCCTCTTTGACTCTCAGTGCCTCAAGTCTCACCACACAGCTGACAGAGAGGACCG	325			

```

Db      238 AGCTCCCTCTTACCTCTCAGCTCAGCTCAAGCTCCACACAGCCTGAGAGAGATGAGCCG 297
QY      326 GACCTGCGGCGACCTGGTGTGCTGTGAACACTCTGCGGCGATCCAGGCGTGCATAGGCA 385
Db      298 GACCTGCGGCGACCTGGTGTGCTGTGAACACTCTGCGGCGATCCAGGCGTGCATAGGCA 357
QY      386 CCGCGCGGCGCGCTGCGCCACTGTGTGCTGTAGCCCACTGCAAGCCCCAGTGTGGCTGCAAC 445
Db      358 CCGCGCGGCGCGCTGCGCCACTGTGTGCTGTAGCCCACTGCAAGCCCCAGTGTGGCTGCAAC 417
QY      446 TTACTGCGCAAGCTCGGAGCGCTGCGCTTGTAGCCTCATGCGCAGCCTCTGTGAGGAGCCTC 505
Db      418 TTACTGCGCAAGCTCGGAGCGCTGCGCTTGTAGCCTCATGCGCAGCCTCTGTGAGGAGCCTC 477
QY      506 AGCCACATTTAGAGGGCTGAGTCAAGCTCCCAACCTTTGGCAGAGAGGGGCCACAGAGC 565
Db      478 AGCCACATTTAGAGGGCTGAGTCAAGCTCCCAACCTTTGGCAGAGAGGGGCCACAGAGC 537
QY      566 CGTAGCATGCGGGGAGCAGCCGCCAGCCTGGGTGCTGCTGGAGCCTGCTGGGCCAGCAGCT 625
Db      538 CGTAGCATGCGGGGAGCAGCCGCCAGCCTGGGTGCTGCTGGAGCCTGCTGGGCCAGCAGCT 597
QY      626 GGCTGCTACTGAGCATGGGCTTGAAGGCGCTTTGAGGATATTACACCTCTATGTAT 685
Db      598 GGCTGCTACTGAGCATGGGCTTGAAGGCGCTTTGAGGATATTACACCTCTATGTAT 657
QY      686 GACAAATGAATTTGGGCGACAGCCTCTGAGGGCGCTCAAAACAGGCCCTAGATGGGCGG 745
Db      658 GACAAATGAATTTGGGCGACAGCCTCTGAGGGCGCTCAAAACAGGCCCTAGATGGGCGG 717
QY      746 GGCAGAGAGAGAGCTCCGAGCTGAGCAGAGGCCGGAATTTGACTCATTTGATGTGCTG 805
Db      718 GGCAGAGAGAGAGCTCCGAGCTGAGCAGAGGCCGGAATTTGACTCATTTGATGTGCTG 777
QY      806 GTGGGCGACACAGGCAATGAGGAGCAGCGGGGCGCGGCGCTGAGCCTCTGCGCTGGAA 865
Db      778 GTGGGCGACACAGGCAATGAGGAGCAGCGGGGCGCGGCGCTGAGCCTCTGCGCTGGAA 837
QY      866 TGGTGTCTGTGATCTGAACCTGAGCTGTGGCTGAGCAGCACTGTCTCGAAAAGACACA 925
Db      838 TGGTGTCTGTGATCTGAACCTGAGCCTGTGGCTGAGCAGCACTGTCTCGAAAAGACACA 897
QY      926 GCTGGCTTCCCTAGTACAGAGAGAGGCTTGGGCCACTTTGGAGAGACAGAAATCTAGTC 985
Db      898 GCTGGCTTCCCTAGTACAGAGAGAGGCTTGGGCCACTTTGGAGAGACAGAAATCTAGTC 957
QY      986 CTGGGCAACTTCACATCCGCTCCCTGCTCAAGGCTGTGCA -CGGGGAGCCCTGGAAATTAC 1044
Db      958 CTGGGCAACTTCACATCCGCTCCCTGCTCAAGGCTGTGCAAGGCGGAGCCTGTGAAATTAC 1017
QY      1045 CCCCTAGTATGAAATGACAGGGTGTGGGAGTGAATTCCTGCGCCCTGGGGGTGATTA 1104
Db      1018 CCCCTAGTATGAAATGACAGGGTGTGGGAGTGAATTCCTGCGCCCTGGGGGTGATTA 1077
QY      1105 GCTTGGGCGTTCCTCTCTCTGTATACGGAAGAGACCCCAATCAGATTTTCAAAATTAAAG 1164
Db      1078 GCTTGGGCGTTCCTCTCTCTGTATACGGAAGAGACCCCAATCAGATTTTCAAAATTAAAG 1137
QY      1165 CCAGTCCCTGGGAATCTCAAAAAAAA 1192
Db      1138 CCAGTCCCTGGGAATCTCAAAAAAAA 1165

```

```

RESULT 7
LOCUS   AY130860          5917 bp    DNA       linear    PRI 22-JUL-2002
DEFINITION Homo sapiens SCDK4-binding protein p34SEI1 (SEI1) gene, complete
ACCESSION AY130860
VERSION   AY130860.1   GI:21914609
KEYWORDS
SOURCE    Homo sapiens (human)

```

```

ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 5917)
AUTHORS Rieder M.J., Braun A.C., Montoya M.A., Chung M.-W., Nguyen C.P.,
Nguyen D.A., Livingston R.J., Foell C.L., Robertson P.D.,
Schackwitz W.S., Sherwood J.K., Wiltrax L.A. and Nickerson D.A.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2002) Genome Sciences, University of Washington,
1705 NE Pacific, Seattle, WA 98195, USA
COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome
Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA
(URL: http://egp.gs.washington.edu).
FEATURES
source
1..5917
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1..274
/rpt_family="Alu"
/rpt_type="dispersed"
170
/frequency="0.09"
/replace="a"
249
/frequency="0.03"
/replace="t"
257
/frequency="0.08"
/replace="a"
267
/frequency="0.04"
/replace="a"
275..402
/rpt_family="Alu"
/rpt_type="dispersed"
562..839
/region="Region not scanned for variation"
593..760
/rpt_family="Alu"
/rpt_type="dispersed"
625
/frequency="0.05"
/replace="a"
782..1108
/rpt_family="Alu"
/rpt_type="dispersed"
977
/frequency="0.05"
/replace="t"
1047
/frequency="0.01"
/replace="a"
1131..1331
/rpt_family="MIR"
/rpt_type="dispersed"
1166
/frequency="0.04"
/replace="t"
1307..1827
/region="Region not scanned for variation"
1410..1524
/rpt_family="MIR"
/rpt_type="dispersed"
1851..5390
/gene="SEI1"
/gene="SEI1"
/product="SCDK4-binding protein p34SEI1"
1887
/gene="SEI1"
/frequency="0.08"
/replace="c"

```

variation	2038	/gene="SEI1"	/frequency="0.44"	/replace="a"
variation	2086	/gene="SEI1"	/frequency="0.01"	/replace="g"
variation	2113	/gene="SEI1"	/frequency="0.01"	/replace="g"
misc_feature	2343	/note="Region not scanned for variation"		
repeat_region	2610	/rpt_family="L2"		
misc_feature	2810	/note="Region not scanned for variation"		
variation	2847	/gene="SEI1"	/frequency="0.09"	/replace="g"
variation	3178	/gene="SEI1"	/frequency="0.07"	/replace="g"
misc_feature	3250	/note="Region not scanned for variation"		
repeat_region	3425	/rpt_family="Mariner"		
repeat_region	3529	/rpt_type="dispersed"		
repeat_region	3683	/rpt_type="dispersed"		
repeat_region	3822	/rpt_type="dispersed"		
misc_feature	4036	/note="Region not scanned for variation"		
repeat_region	4144	/rpt_family="Alu"		
repeat_region	4303	/rpt_type="dispersed"		
repeat_region	4393	/note="Region not scanned for variation"		
repeat_region	4788	/rpt_family="Alu"		
repeat_region	4812	/rpt_type="dispersed"		
repeat_region	4901	/rpt_type="dispersed"		
repeat_region	5013	/rpt_type="dispersed"		
repeat_region	5125	/rpt_type="dispersed"		
repeat_region	5237	/rpt_type="dispersed"		
repeat_region	5349	/rpt_type="dispersed"		
repeat_region	5461	/rpt_type="dispersed"		
repeat_region	5573	/rpt_type="dispersed"		
repeat_region	5685	/rpt_type="dispersed"		
repeat_region	5797	/rpt_type="dispersed"		
repeat_region	5909	/rpt_type="dispersed"		
repeat_region	6021	/rpt_type="dispersed"		
repeat_region	6133	/rpt_type="dispersed"		
repeat_region	6245	/rpt_type="dispersed"		
repeat_region	6357	/rpt_type="dispersed"		
repeat_region	6469	/rpt_type="dispersed"		
repeat_region	6581	/rpt_type="dispersed"		
repeat_region	6693	/rpt_type="dispersed"		
repeat_region	6805	/rpt_type="dispersed"		
repeat_region	6917	/rpt_type="dispersed"		
repeat_region	7029	/rpt_type="dispersed"		
repeat_region	7141	/rpt_type="dispersed"		
repeat_region	7253	/rpt_type="dispersed"		
repeat_region	7365	/rpt_type="dispersed"		
repeat_region	7477	/rpt_type="dispersed"		
repeat_region	7589	/rpt_type="dispersed"		
repeat_region	7701	/rpt_type="dispersed"		
repeat_region	7813	/rpt_type="dispersed"		
repeat_region	7925	/rpt_type="dispersed"		
repeat_region	8037	/rpt_type="dispersed"		
repeat_region	8149	/rpt_type="dispersed"		
repeat_region	8261	/rpt_type="dispersed"		
repeat_region	8373	/rpt_type="dispersed"		
repeat_region	8485	/rpt_type="dispersed"		
repeat_region	8597	/rpt_type="dispersed"		
repeat_region	8709	/rpt_type="dispersed"		
repeat_region	8821	/rpt_type="dispersed"		
repeat_region	8933	/rpt_type="dispersed"		
repeat_region	9045	/rpt_type="dispersed"		
repeat_region	9157	/rpt_type="dispersed"		
repeat_region	9269	/rpt_type="dispersed"		
repeat_region	9381	/rpt_type="dispersed"		
repeat_region	9493	/rpt_type="dispersed"		
repeat_region	9605	/rpt_type="dispersed"		
repeat_region	9717	/rpt_type="dispersed"		
repeat_region	9829	/rpt_type="dispersed"		
repeat_region	9941	/rpt_type="dispersed"		
repeat_region	10053	/rpt_type="dispersed"		
repeat_region	10165	/rpt_type="dispersed"		
repeat_region	10277	/rpt_type="dispersed"		
repeat_region	10389	/rpt_type="dispersed"		
repeat_region	10501	/rpt_type="dispersed"		
repeat_region	10613	/rpt_type="dispersed"		
repeat_region	10725	/rpt_type="dispersed"		
repeat_region	10837	/rpt_type="dispersed"		
repeat_region	10949	/rpt_type="dispersed"		
repeat_region	11061	/rpt_type="dispersed"		
repeat_region	11173	/rpt_type="dispersed"		
repeat_region	11285	/rpt_type="dispersed"		
repeat_region	11397	/rpt_type="dispersed"		
repeat_region	11509	/rpt_type="dispersed"		
repeat_region	11621	/rpt_type="dispersed"		
repeat_region	11733	/rpt_type="dispersed"		
repeat_region	11845	/rpt_type="dispersed"		
repeat_region	11957	/rpt_type="dispersed"		
repeat_region	12069	/rpt_type="dispersed"		
repeat_region	12181	/rpt_type="dispersed"		
repeat_region	12293	/rpt_type="dispersed"		
repeat_region	12405	/rpt_type="dispersed"		
repeat_region	12517	/rpt_type="dispersed"		
repeat_region	12629	/rpt_type="dispersed"		
repeat_region	12741	/rpt_type="dispersed"		
repeat_region	12853	/rpt_type="dispersed"		
repeat_region	12965	/rpt_type="dispersed"		
repeat_region	13077	/rpt_type="dispersed"		
repeat_region	13189	/rpt_type="dispersed"		
repeat_region	13301	/rpt_type="dispersed"		
repeat_region	13413	/rpt_type="dispersed"		
repeat_region	13525	/rpt_type="dispersed"		
repeat_region	13637	/rpt_type="dispersed"		
repeat_region	13749	/rpt_type="dispersed"		
repeat_region	13861	/rpt_type="dispersed"		
repeat_region	13973	/rpt_type="dispersed"		
repeat_region	14085	/rpt_type="dispersed"		
repeat_region	14197	/rpt_type="dispersed"		

[illegible]

Db	117961	GGTCATACCTTGGGCGTTCCTTCCTCCTGATACGGGAAGACCCCAATCAGATTTTTCAA	117902
Qy	1158	ATTAAGCCAGCTCTGGGAATCTCATA	1185
Db	117901	ATTAAAGCCAGCTCTGGGAATCTCATA	117874
RESULT 9	AX472956	5746 bp	DNA
LOCUS	AX472956	Sequence 5 from Patent WO0218576.	linear
DEFINITION	AX472956		
ACCESSION	AX472956.1	GI:22207743	
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Chen, S. Y., Macina, R. A., Sun, Y. and Recipon, H.		
TITLE	Compositions and methods relating to lung specific genes		
JOURNAL	Patent: WO 0218576-A 5 07-MAR-2002;		
Diadexus, Inc. (US)			
FEATURES			
Source	1..5746	Location/Qualifiers	
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	1141 a 1660 c 1980 g 965 t		
ORIGIN			
Query Match	65.3%;	Score 777.8;	DB 6;
Best Local Similarity	99.7%;	Pred. No. 1.1e-161;	
Matches 779;	Conservative 0;	Mismatches 2;	Indels 0; gaps 0;
Qy	26	CGCGACCCAGGGGGGTTTCGGAGGAGCAGCCAAACGAGTGCCTGGCCGCTTTC	85
Db	1	CGCGACCCAGGGGGGTTTCGGAGGAGCAGCCAAACGAGTGCCTGGCCGCTTTC	60
Qy	86	CTGATTTGGTTGGGTGGCTACCTCTTCTGATTTGGTCCCTAGTGAAGATGCTG	145
Db	61	CTGATTTGGTTGGGTGGCTACCTCTTCTGATTTGGTCCCTAGTGAAGATGCTG	120
Qy	146	AGCAAGGTCGGAAGCGGAACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	205
Db	121	AGCAAGGTCGGAAGCGGAACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	180
Qy	206	TCTGTGTGGCTAGATCTTGGCCACACGAGCGGTGGACAGGACCCCGGCTGTG	265
Db	181	TCTGTGTGGCTAGATCTTGGCCACACGAGCGGTGGACAGGACCCCGGCTGTG	240
Qy	266	AGCTCCCTCTTTGACCTCTCAAGTGTCAAGCTTCACACACAGCTGTGACAGTGA	325
Db	241	AGCTCCCTCTTTGACCTCTCAAGTGTCAAGCTTCACACACAGCTGTGACAGTGA	300
Qy	326	GACCTGGCGACCTGGTGTCTGTGTAACACTTGGCGGCGATCCAGGCTTCATGGCA	385
Db	301	GACCTGGCGACCTGGTGTCTGTGTAACACTTGGCGGCGATCCAGGCTTCATGGCA	360
Qy	386	CCGCGGGCTGCCACCTGTGTGAGCCACCTGACGCCACAGTGTGCTGCAAC	445
Db	361	CCGCGGGCTGCCACCTGTGTGAGCCACCTGACGCCACAGTGTGCTGCAAC	420
Qy	446	TTACTGGCAACTGGAGGTGCCCTTTCAAGCTTCATAGGCGCAGCTCTTGGAGACCTC	505
Db	421	TTACTGGCAACTGGAGGTGCCCTTTCAAGCTTCATAGGCGCAGCTCTTGGAGACCTC	480
Qy	506	AGCACAATTGAGGCGCTGATCAGGCTCCCAACCTTGGCGAGAGAGGCGCACAGGC	565
Db	481	AGCACAATTGAGGCGCTGATCAGGCTCCCAACCTTGGCGAGAGAGGCGCACAGGC	540
Qy	566	CGTAGCATGGGGGAGACGCGCCACCTGTGGGTGCTTTGACCTGTGGGCCAGCACT	625

Db	541	CGTAGCATGGGGGAGCAGCCGCCAGCCTGGGTGGCTTGGACCTGCTGGCCCAAGCACT	600
QY	626	GGCTGTCTACTGAGCAGATGGGCTTGAGGGCCTGTTTGAGATATTGACACCTATGTAT	685
Db	601	GGCTGTCTACTGAGCAGATGGGCTTGAGGGCCTGTTTGAGATATTGACACCTATGTAT	660
QY	686	GACATGAACTTTTGGGGACCAAGCCCTGATAGGGCCCAACAAGGCCCGGAGATGGGGCG	745
Db	661	GACATGAACTTTTGGGGACCAAGCCCTCTAGGGCCTCAACAAGGCCCTGAGAGATGGGGCG	720
QY	746	GGCAAGAGAGAGAGCTCCGAGCTGTGACAGAGGCCGGAATTGGACTTACCTCATGGATGGCTG	805
Db	721	GGCAAGAGAGAGAGCTCCGAGCTGTGACAGAGGCCGGAATTGGACTTACCTCATGGATGGCTG	780
QY	806	G	806
Db	781	G	781
RESULT 10			
LOCUS	HSJ323696		
DEFINITION	Homo sapiens genomic sequence surrounding NotI site, clone		
ACCESSION	NB5-544S.		
VERSION	AJ323696		
KEYWORDS	AJ323696.1 GI:15868075		
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 978)		
AUTHORS	Kusenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Krasha,S.M., Podowski,R.M., Matushkin,Y.G., Granchandani,A., Muravenko,O.V., Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashba,V.I., Kiselev,I.L., Wasserman,N., Wahlestedt,C. and Zabarovsky,E.R.		
TITLE	NotI flanking sequences: a tool for gene discovery and verification of the human genome		
JOURNAL	Nucleic Acids Res. 30 (14), 3163-3170 (2002)		
MEDLINE	12131767		
PUBMED	12136098		
REFERENCE	2 (bases 1 to 978)		
AUTHORS	Zabarovsky,E.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-May-2001) Microbiology and Tumorbiology Centre, Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77, Sweden		
FEATURES			
source	location/Qualifiers		
	1..978		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/clone="NB5-544S"		
BASE COUNT	188 a 300 c 299 g 188 t 3 others		
ORIGIN			
Query Match	61.9%; Score 738; DB 9; Length 978;		
Best Local Similarity	96.8%; Pred. No. 9e-153;		
Matches 763; Conservative	0; Mismatches 7; Indels 2; Gaps 2;		
QY	200	GTGCACATCCCTGGTGGCTAGATCTGGCCACAGCGGTGGGCACAGACCCCGGCGCTG	259
Db	1	GTGCACATCCCTGGTGGCTAGATCTGGCCACAGCGGTGGGCACAGACCCCGGCGCTG	60
QY	260	GCCCTAGCTCCCTCTTTGACCTCTCAGTGTCCAGGCTCCACACAGCCTGACAGAGT	319
Db	61	GCCCTAGCTCCCTCTTTGACCTCTCAGTGTCCAGGCTCCACACAGCCTGACAGAGT	120
QY	320	GAGCGGACCTCGGGGACACCTGGTCTGTGTGAACACTCTGGGGCGCATCAGAGCTCC	379
Db	121	GAGCGGACCTCGGGGACACCTGGTCTGTGTGAACACTCTGGGGCGCATCAGAGCTCC	180

OY		380	ATGGACACC	GCGGCGCCCTGGCACCTTGCCGTACCCACAGCCGGAAGCCCCAGTGGCT	439
Dd		181	ATGGCACC	C CGGCGTCGCCCTGGCACCTTGCCGTACCCACAGCCGGAAGCCCCAGTGGCT	240
OY		440	GACAACCT	TACTTAGGCAAGCTCGGAGCCTCCCTTTAGCCTCCATGGCAGCCTCTGGAG	499
Dd		241	GACAACCT	TACTTAGGCAAGCTCGGAGCCTCCCTTTAGCCTCCATGGCAGCCTCTGGAG	300
OY		500	GACCTAGC	CACACATTGAGGGCCTGAGTAGGCTCCCAACCCCTTGGCAGACGAGGGCCA	559
Dd		301	GACCTAGC	CACACATTGAGGGCCTGAGTAGGCTCCCAACCCCTTGGCAGACGAGGGCCA	360
OY		560	CCAGGCCT	GTAGCATGCGGGGGAGACAGCCGCCAAGCCCTGGCTTGAGGAACTGCTGGGCCCA	619
Dd		361	CCAGGCCT	GTAGCATGCGGGGGAGACAGCCGCCAAGCCCTGGCTTGAGGAACTGCTGGGCCCA	420
OY		620	GCCATGGC	TGCTGTCTAGAGCATGGGCTTGAGGGCCTGTTTGAAGATATTGACACCTCT	679
Dd		421	GCCATGGC	TGCTGTCTAGAGCATGGGCTTGAGGGCCTGTTTGAAGATATTGACACCTCT	480
OY		680	ATGTATGA	CAATGAACCTTTGGGACACAGCCTCTGAGGGCCTCAAACAGGCCCTGAGGAT	739
Dd		481	ATGTATGA	CAATGAACCTTTGGGACACAGCCTCTGAGGGCCTCAAACAGGCCCTGAGGAT	540
OY		740	GGGGCGGG	CGCAAGGAGAGCTCCGAGCTGGACAGAGGCCGAATTGACTACTATGAT	799
Dd		541	GGGGCGGG	CGCAAGGAGAGCTCCGAGCTGGACAGAGGCCGAATTGACTACTATGAT	600
OY		800	GTGCTGTG	GGGGCACACAGGACCTGGAGGAGCGCGGGGCCAAGGCGGTGAGCCCTGCTG	859
Dd		601	GTGCTGTG	GGGGCACACAGGACCTGGAGGAGCGCGGGGCCAAGGCGGTGAGCCCTGCTG	660
OY		860	CTGGAATG	TGTGTGTGATCTGAAGTAAGCCTGCTGGCTGGACCAATGCTCTGAAAA	919
Dd		661	CTGGAATG	TGTGTGTGATCTGAAGTAAGCCTGCTGGCTGGACCAATGCTCTGAAAA	720
OY		920	GACACAG-	-CTGGCTTCCCTAGTACAGAGA- GGGCTTGGGCCACTTTGGA	969
Dd		721	GACACAGN	CTGCTTCCCTAGTACAGAGACAGGGGCTTGGGCCACTTTGGA	772
RESULT 11					
AF366401			1247 bp	mRNA	linear ROD 13-MAY-2001
LOCUS					
DEFINITION	Mus musculus TRIP-Brl mRNA, complete cds.				
ACCESSION	AF366401				
VERSION	AF366401.1	GI:14029831			
KEYWORDS	.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 1247)				
AUTHORS	Hsu,S.I., Yang,C.M., Sim,K.G., Hentschel,D.M., O'Leary,E. and Bonventre,J.V.				
	TRIP-Brl: a novel family of PHD zinc finger- and bromodomain-interacting proteins that regulate the transcriptional activity of E2F-1/DP-1				
JOURNAL	EMBO J. 20 (9), 2273-2285 (2001)				
	21231173				
MEDLINE	11331592				
PUBMED	2 (bases 1 to 1247)				
REFERENCE	Hsu,S.I. and Bonventre,J.V.				
AUTHORS	Direct Submission				
JOURNAL	Submitted (29-MAR-2001) Dept. of Medicine, National University Hospital, 5 Lower Kent Ridge Road 119074, Singapore				
FEATURES	Location/Qualifiers				
source	1..1247				
	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/db_xref="dbEST:W82777"				
	/db_xref="taxon:10090"				
CDS	173..883				

		/note="TRIP-B1, transcriptional regulator interacting with the PHD-Dromodomain, similar to Homo sapiens TRIP-B1 (Seq-1)"		
		/codon_start=1		
		/product="TRIP-B1"		
		/protein_id="AAK52830.1"		
		/db_xref="GI:14029832"		
		/translation="MLSGLRKRREETMEALSVDSCLMDPSHPVAQNPPTVASSS		
		LFDSVVLHSHLSRQSEPDHLVLYVWTLRRIQASMPAPVLPPEIOPAPASVDS		
		LLASDADLSASMSLEIDNLHEDLNAPPOADGPGGSGISINLIDALDILG		
		ATGCLDGLGLEDIEDITSMYDSBLMLPABEGLKPEPENPABEPEDEAEIDLY		
		MDVYVQALRRPGRGR"		
BASE COUNT	276 a	352 c	372 g	247 t
ORIGIN				
Query Match	56.7%	Score 675.6:	DB 10:	Length 1247:
Best Local Similarity	76.4%	Pred. No. 5.6e-139:		
Matches	931:	Conservative	0:	Mismatches 254; Indels 33; Gaps 7;
QY	1	GTGACCCGGAGTCAGAACTGCGTCTCGGACCCAGCGCGGGTTCGCGAGACAGCCCA	60	
Db	31	GTGCGCCGGAGTCAGAACTGCGGCTTACCGTCCAGAGTGTGCGTGTGCGAGCAAG	90	
QY	61	CAAGCGATGCTGCGCGCCCGCTTTCCTATTGCTTGTTGGTGGCTACTCTTTCGTTCTGA	120	
Db	91	CTAGGAGTCACGCGCACCGCCCTTCCTATTGCGCTGTGCGCTCCCTCTTCCTTCGTA	150	
QY	121	TTGGCCGCTATGTA--GCAAGATGCTGAGCAAGGCTGTGAAGGGGAACGGGAGGAGGA	177	
Db	151	TTGGCCGCGAGTGTCTGTAGAGATGTGAGCAAGGCTGTGAAGCGCAAGCGGAGGAGGA	210	
QY	178	GGAGGAGAGAAACCTCTGCGCAGTTCGATCCTGCTGAGCTATGCTCGACAGCAGCGGT	237	
Db	211	GGAGGAGATGGAAGCCCTCTCCGTGAGTCTGTCTGTGTGATTCGAGCACCAGCGCGAGT	270	
QY	238	GGCAAGGCAACCCCGCGCGCTGAGCTTCCCTTTGACCTCTTCAGTGCCTCAAGCT	297	
Db	271	GGCGCAGACCCCGCGGAGCGTGTGCTTCACTCCCTCTTGAACCTTCAGTGTCAAGCT	330	
QY	298	CCACCAACGCTGTGAGCAGATGAGCGGAGCACTGCGCACTGCTGTGCTGTGAGAAC	357	
Db	331	CCACCAACGCTGTGAGCAGCAGCAGCGGAGCCTCGGACCTGTGCTGTGAGAAC	390	
QY	358	TCTGCGCGCATCCAGAGCGCTCCATGAGCAACCGCGCGCTCCCTGCACTTGCGCTAGCGCC	417	
Db	391	ACTAGAGCGCATCCAGAGCGCTCCATGAGCAACCGCACTGTCTCTGCGCGCTGAGCCATCA	450	
QY	418	ACCTGCAAGCCCAAGTGTGCTGAACAATTACTGGCAAGCTGGACGCTGCGCTTACG	477	
Db	451	GCCCCAGACCCCTTCTGTGTGCGCAGCAGGCTTCTGGTAGCTCAATGCTGTGCGCTCAGC	510	
QY	478	CTCATGCGCAGCCCTCGTGGAGAGCACTGAGCAATGAGGCGCTGATCAGGCTCCCA	537	
Db	511	CTCATGCGCAGCCCTCGTGGAGAGCTCAACAACATTTGAGGAGCACTGGAACGCGCCCA	570	
QY	538	ACCCCTGGCAGACAGGGGCCACCAAGCCGTAGCATGCGGGGAGCAGCAGCCAGCCTGGG	597	
Db	571	ACCCCAAGCAGATAGAGGGGCTCCAGGCGGCTCATGCGGGGAGATCTCACCAACCTGGG	630	
QY	598	TGCGTTGGACCTGCTGGGCGCCAGCCACTGCGCTGTCTACTGTGACAGATGGGCTTGAAGGCT	657	
Db	631	TGCGTTGGACCTGCTAGGCGCCAGCCAGCAGCTGTCTGTCTGTGACATGAGTGAAGGCGCT	690	
QY	658	GTTTGAGAGATTTACACACTCTATGTATGACAAATGAACCTTGGGACACAGCTCTGAGGG	717	
Db	691	GTTTGAGAGATTCATACCTTCATGATGACGACAGTGAACCTTGGTTTACAGGCTCTGAGGG	750	
QY	718	CCTCAAAACAGAGCCCTAGAGATGGGCGCGGCAAGAGAAAGCTCCGGAGCTGTGACAGGCG	777	
Db	751	TCTCAAGCCGCGCCTGAGATGGCCCAAGCAGAGAGAGAGCTCTCAAGATTTGATGAGGC	810	
QY	778	CGAATTGAGCTACCTCATGTGATGTGCTGTGGGACACAGGCACTGTGAGGAGCAGCGCGGG	837	

Db	811	GGAGTGTGGACTACCTCACTGATGTACTAGTAGGACAACAGGCACCTGGAAAAGCCACCAGG	870
QY	838	GCCAGGGGCGCTGAGACCCCTCG--TCTGGAAATGGTTGTCTGGTAATCTGAACCTGAGCTTCT	895
Db	871	GCCCTGGGGGCGCTGACCCCTTAGACCACCGAGGAATAATGTACTTGGATCTGGTTGAGACCTTGGT	930
QY	896	GGCTGAGACCACTCTCCTCGAAAGAGACACAGCTGGCTCCCTAGTACAGAGAACAGGGCT	955
Db	931	GGTGCACACCAACCCACTTTGACACGACACACAGGTGGCGTCCC--CTACACAGAGAGAGCC	987
QY	956	TGGGCCCACTTTGGAGAGACAGAAATCTAGTCTCTGGGCAACTTCACATCCGTCCTCTGTCT	1015
Db	988	AGGGTCACTTTGGAGAGACTGTACTGAGTCCCTGGGG-AACTTCACATCCATCCACTCTGCTC	1046
QY	1016	CAGGGCTGGCAGGGGGGAGCTGGAAATTAACCCCTAGTATGATGATGACAGGGTCT-----	1069
Db	1047	---TGACCAGTGAAGAGAGTTGGAAATGAGCCCCAGATGTGTAAATATGACAGGGGCCCACTGA	1103
QY	1070	-----TGGTGGGAGCAATTAATCCCTGGGCCCTGGGGGCTGATAGCTTGGGCTGTWCC--	1118
Db	1104	CAGTAATAGTGTTTGGGGCCAGCGTGGGGTTAACTGGGATTAATAGTTTGGGCTATTCOCAT	1163
QY	1119	----TTCCTGATATCGGGAGAGACCCCAATCAGATTTTTCAAATTAAGCCAGTCTGTG	1174
Db	1164	CCCATTAATCCCGACTGTGGGAGAGAGACCTCAATATTTTGTAAATTAAGAACACAGCTTTGG	1223
QY	1175	GAATCTCAAAAAAAAAA	1192
Db	1224	GAATTAATAAAAAAAAAA	1241

RESULT	12		
AX537459			
LOCUS	AX537459	5869 bp	DNA
DEFINITION	Sequence 30 from Patent WO020709.		linear
ACCESSION	AX537459		
VERSION	AX537459.1	GI:25269253	
KEYWORDS			
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens (human)		

REFERENCE
AUTHORS
1. Lal, P. G., Baughn, M. R., Yao, M. G., Malla, N. K., Elliot, V. S., Xu, Y., Hochella, C. D., Yue, H., Ding, L., Gietzen, K. J., Ison, C. H., Lu, D. A., Hochella, A. J., Ghadai, A. R., Changavelu, K. S., Sanjavalu, M. M., Tang, Y. T., Runkumar, J., Griffin, T. A., Swaraker, A., Azimal, Y., Sapsstee, S. K., Burford, N., Lee, E. A., Lu, Y., Tian, D. K. and Matquais, U. P.

TITLE
Molecules for disease detection and treatment
JOURNAL
Patents: WO 02070709; A 30 12-SEP-2002;

FEATURES	SOURCE
Location/Qualifiers	1..5869
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/note="Incyte ID No: 290316CB1"	
1191 a 1682 c 1995 g 1001 t	

	Query Match	Score	DB	Length
Best Local Similarity	99.9%	Pred. No. 3.1e-137		
Matches	668	Conservative	0	Mismatches 1; Indels 0; Gaps 0;
Oy	138	AGATGCGAGCAAGGGCTGTGAAGCGGAAACGGGAGAGGAGGAGAGAGAACTCTG	197	
Db	237	AGATGCTGACCAAGCGTCTGAACGGGAAACGGGAGAGGAGAGAGACAACTCTG	296	
Oy	198	CAGTCGACTCTGTGGCTAGACTCTGGCCACGACGCGGTGGCAOAGGACACCCCGGCG	257	
Db	297	CAGTCGACTCTGTGGCTAGACTCTGGCCACACAGCGGTGGCCACAGGACACCCCGGCG	356	

OY	258	TGGGCTGAGTCCCTCTTTTACCTGACCTGCAGAGCTCCAAAGCTCCACACAGCCTGGAGAGA	312
Db	357	TGGCCTTAGTCCCTCTTTTACCTCTCAGTGTCTCAAGCTCCACCAAGCCTGCAGCAGA	416
OY	318	GTGACCGGACCTCGGGACCTGGTGTGTGTGTGTGAACACTCTGCGGGCGCATCGAGCGT	377
Db	417	GTGAGCCGACCTCGGGACCTGGTGTGTGTGTGTGAACACTCTGCGGGCGCATCGAGCGT	476
OY	378	CCATGGCACCCGGGGCTGGCCCTGGCACCTGTGCTTACCCACCTGCGAGCCGCCAGTGTGG	437
Db	477	CCATGGCACCCGGGGCTGGCCCTGGCACCTGTGCTTACCCACCTGCGAGCCGCCAGTGTGG	536
OY	438	CTGACAACTTACTGGCAAGCTCGGACGCTGGCCCTTTCAGCCTCCATGAGCGACGCTCGTGG	497
Db	537	CTGACAACTTACTGGCAAGCTCGGACGCTGGCCCTTTCAGCCTCCATGAGCGACGCTCGTGG	596
OY	498	AGGACCTGACCCACATTTGAGGGGCTGAGTCAAGCTCCCAACCTTTGGCAGACGAGGGGC	557
Db	597	AGGACCTGACCCACATTTGAGGGGCTGAGTCAAGCTCCCAACCTTTGGCAGACGAGGGGC	656
OY	558	CACAGAGCCCTAGCATCGGGGGAGCAGGCCACGCTGAGGTCGCTTGGACCTGCTGGGGC	617
Db	657	CACAGAGCCCTAGCATCGGGGGAGCAGGCCACGCTGAGGTCGCTTGGACCTGCTGGGGC	716
OY	618	CAGCCACTGCGTGTCTACTGGACGATGGGCTTGAAGGCGCTGTTTGAAGATATTGACACT	677
Db	717	CAGCCACTGCGTGTGTCTACTGGACGATGGGCTTGAAGGCGCTGTTTGAAGATATTGACACT	776
OY	678	CTATGTATGACAAATGAAGACTTTGGGCACACGCTCTGAGGGGCGTCAAAACAGAGCCCTAGG	737
Db	777	CTATGTATGACAAATGAAGACTTTGGGCACACGCTCTGAGGGGCGTCAAAACAGAGCCCTAGG	836
OY	738	ATGGGCGCGGCAAGAGAGAGCTCCGGAGCTGAGCAGAGCGAATTTGGACTACTCTCATGG	797
Db	837	ATGGGCGCGGCAAGAGAGAGCTCCGGAGCTGAGCAGAGCGAATTTGGACTACTCTCATGG	896
OY	798	ATGTGCTGG 806	
Db	897	ATGTGCTGG 905	

RESULT	13
LOCUS	BC016077
DEFINITION	BC016077 1165 bp mRNA linear ROD 16-Apr-2003
ACCESSION	Mus musculus RIKEN cDNA 1110032C13 gene, mRNA (cDNA clone MGC:27573
VERSION	IMAGE:486318), complete cds.
KEYWORDS	BC016077 BC016077 . GI:16359221
SOURCE	MGC.
ORGANISM	Mus musculus (house mouse)
REFERENCE	Mus musculus
AUTHORS	Euharyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1165) Straussberg,R.L., Feinold,E.A., Grosse,L.H., Dere,J.G.,

Altschul, S.F., Zeeberg, B., Buelow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavante, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Wooley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Vittalton, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A.C., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shvachenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Gilmwood, J., Schmutz, J., Myers, R.M., Butlerfield, I.S., Krzywinski, M.I., Skalska, U., Smallus, D.E., Schermer, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Df						1135	CITGGGAATATAAAAAAAAAA	1158	
RESULT 14									
LOCUS	BD124736								
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.								
ACCESSION	BD124736								
VERSION	JP 2002017375-A/167.								
KEYWORDS	Homo sapiens (human)								
SOURCE	Homo sapiens								
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 696) Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.								
REFERENCE	Primer for synthesizing full-length cDNA and use thereof								
AUTHORS	Patent: JP 2002017375-A 167 22-JAN-2002;								
TITLE	HELIX RESEARCH INSTITUTE								
JOURNAL	OS Homo sapiens (human)								
COMMENT	PN 22 JAN-2002 A/167 PD 22 JAN-2002 PE 07-JUL-2000 JP 2000235172 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO PI PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUOT OTSUKI,HISASHI KOGA PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC 10, C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC PC Primer for synthesizing full-length cDNA and use thereof FH key Location/Qualifiers FT source 1..696 FT /organism='Homo sapiens (human)'. FEATURES source Location/Qualifiers 1..696 /organism='Homo sapiens' /mol_type='genomic DNA' /db_xref='taxon:9606'								
BASE COUNT	120 a 221 c 220 g 132 t 3 others								
ORIGIN									
Query Match	53.5%; Score 637.2; DB 6; Length 696;								
Best Local Similarity	98.1%; Pred. No. 1.9e-130;								
Matches	664; Conservative 0; Mismatches 11; Indels 2; Gaps 2;								
OY	1 GTGAGCCGGAGTCAGAAGTGCCTCGGTCCGAGCCCCAGCGCGGTTCCGGAGACGCCAA 60								
Dd	21 GTGAGCCGGAGTCAGAAGTGCCTCGGTCCGAGCCCCAGCGCGGTTCCGGAGACGCCAA 80								
OY	61 CAACGATGTCGGCCCGCGCGTTCCTCGATTGGTGTTGGGGGCTACTCTTCTGTCA 120								
Dd	81 CAACGATGTCGGCCCGCGCGTTCCTCGATTGGTGTTGGGGGCTACTCTTCTGTCA 140								
OY	121 TTGGCCGCTAGTAGACAAGATGCTGAGCAAGGTCCTGAAGCGAAAACGGAGAGAGAGA 180								
Dd	141 TTGGCCGCTAGTAGACAAGATGCTGAGCAAGGTCCTGAAGCGAAAACGGAGAGAGAGA 200								
OY	181 GGAGAAAGAACCTCTGGGAGTCGAGCATCTGGTGGCTGAGATCTMGGCCACGACGAGGTCGC 240								
Dd	201 GGAGAAAGAACCTCTGGGAGTCGAGCATCTGGTGGCTGAGATCTMGGCCACGACGAGGTCGC 260								
OY	241 ACAGGCAACCCCGCGCGCTGACCTTACTCTCTTGGACTCTCAGTGTCAAGCTCCA 300								
Dd	261 ACAGGCAACCCCGCGCGCTGACCTTACTCTCTTGGACTCTCAGTGTCAAGCTCCA 320								
OY	301 CCACAGGCTGAGCAGAGTACCAGCCGGAGCTBGGCACCTGGTGCTGTAAGACTCT 360								

Db	321	CCACAGCGCTGCAGACGAGTGGAGCCGAGCACTGGGCGACCTGGTGTCTGGTGTGAACA	CTCT	380
QY	361	GGGGGGCATCCACAGCGGTCATGGACCCCGCGGCTGCTCCACCTGTGGCTTGA	AGCCAC	420
Db	381	GGGGGGCATTCACAGGGGTCATGGACCCCGCGGCTGCTCCACCTGTGGCTTGA	AGCCAC	440
QY	421	TGCACGCCCCAGTGTGGCTGACACCTTACTTGGCAAGCTGGAGCGCTGGCCCTT	CAGCTC	480
Db	441	TGCACGCCCCAGTGTGGCTGACACCTTACTTGGCAAGCTGGAGCGCTGGCCCTT	CAGCTC	500
QY	481	CATGGCCAGCGCTCTCTGGAGAGCCTCAGCCACATTTGAGGGCCCTGAGTCAG	AGCTCCCAAC	540
Db	501	CATGGCCAGCGCTCTCTGGAGAGCCTCAGCCACATTTGAGGGCCCTGAGTCAG	AGCTCCCAAC	560
QY	541	CTTGGGCACAGAGGGGGCCACACAGGCGCTTGGCATTCGGGGAGAGACAGCC	CCCAACCTGGGTGC	600
Db	561	CTTGGGCACAGAGGGGGCCACACAGGCGCTTGGCATTCGGGGAGAGAGCGCC	CCCAACCTGGGTGC	620
QY	601	CTTGGACCTGCTGGGGCCACAGCCACTGTGCTTACTGTGACGATGGGCTTGA	GGGCGCTGT	660
Db	621	CTTGGACCTGCTGGGGCCACAGCCACTGTGCTTACTGTGACGATGGGCTTGA	GGGCGCTGT	678
QY	661	TGAGCATATTGACACCT	677	
Db	679	TGAGCATTTTGGCCCT	695	
RESULT 15				
LOCUS	BD126760	696 bp	DNA	linear
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.			
ACCESSION	BD126760			
VERSION	BD126760.1	GI:23221705		
KEYWORDS	JP 2002017375-A/2191.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.			
TITLE	Primer for synthesizing full-length cDNA and use thereof			
JOURNAL	Patent: JP 2002017375-A	2191	22-JAN-2002.	
COMMENT	HELIX RESEARCH INSTITUTE OS Homo sapiens (human) PN JP 2002017375-A/2191 PD 22-JAN-2002 PF 07-JUL-2000 JP 2000253172 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO PI ISHII, PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUJI OTSUKI,HISASHI KOGA			
FEATURES				
SOURCE	location/Qualifiers			
	1..696			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:9606"			
BASE COUNT	120 a	221 c	220 g	132 t
ORIGIN				
	Query Match	53.5%	Score 637.2;	DB 6; Length 696;
	Best Local Similarity	98.1%;	Prod. No. 1.9e-130;	
	Matches 664; Conservative	0;	Mismatches 11;	Indels 2;
			Gaps	2;

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)